



**HAL**  
open science

# Influence of the genomic context on enhancer-promoter interactions during early *Drosophila* embryogenesis

Alexia Grasso

► **To cite this version:**

Alexia Grasso. Influence of the genomic context on enhancer-promoter interactions during early *Drosophila* embryogenesis. Molecular biology. Université de Lyon, 2021. English. NNT : 2021LY-SEN053 . tel-04416443

**HAL Id: tel-04416443**

**<https://theses.hal.science/tel-04416443>**

Submitted on 25 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Numéro National de Thèse : 2021LYSEN053

**THESE DE DOCTORAT DE L'UNIVERSITE DE LYON**  
opérée par  
**l'Ecole Normale Supérieure de Lyon**

**Ecole Doctorale N° 340**  
**Biologie Moléculaire, Intégrative et Cellulaire (BMIC)**

**Discipline** : Sciences de la Vie

Soutenue publiquement le 18/10/2021, par :

**Alexia GRASSO**

---

**Influence of the genomic context on  
enhancer-promoter interactions during  
early *Drosophila* embryogenesis**

**Influence du contexte génomique  
sur les interactions enhancer-promoteur au cours de  
l'embryogenèse précoce  
de la Drosophile**

---

Devant le jury composé de :

Dr GIORGETTI Luca, Chef d'équipe, FMI/Basel	Rapporteur
Dr LAGHA Mounia, Directrice de recherche, IGMM/Montpellier	Rapporteuse
Dr OUDELAAR Marieke, Chef d'équipe, Max Planck Institute/Gottingen	Examinatrice
Dr SCHNORRER Frank, Directeur de recherche, IBDM/Marseille	Examineur
Dr AVEROF Michalis, Directeur de recherche, IGFL/Lyon	Examineur
Dr GHAVI-HELM Yad, Chargée de recherche HDR, ENS de Lyon	Directrice de thèse





Influence of the genomic context  
on enhancer-promoter interactions  
during early *Drosophila* embryogenesis



# Acknowledgments

Here we are, almost 4 years since the beginning of my PhD. It is been a long journey full of joy, excitement, success, surprise, satisfaction but also full of mistakes and frustrations that equally have taught me so much.

I have to thank a lot of persons, starting with my PhD supervisor, **Yad Ghavi-Helm**. When we first met in Germany during my Master internship at EMBL I knew I would learn a lot from you. Thank you for choosing me as your first PhD student, thank you for trusting me in this project, for the useful guidance and advice, but above all thank you for always being on my side even in difficult times.

I would like to thank all the current and former members of the lab, who helped me a lot during the process of the project, in particular, thank you to **Hélène Tarayre, Séverine Vincent, Deevitha Balasubramanian and Charlotte Moretti**. A special gratitude goes to **Isabelle Stevant** for her presence and support in scientific discussions and during my thesis writing.

Thanks to my committee members **Kiran Pasmanabhan, Cedric Vaillant, Frank Schnorrer**, which during these 4 years have given me important advices for the progress of the project. I also thank the members of the jury **Frank Schnorrer, Michalis Averof, Marieke Oudelaar, Luca Giorgetti and Mounia Lagha** who agreed to evaluate this work.

I would like to thank **Vitoria Tobias Santos, Aidamalia Vargas-Lowman** and **Cagri Bahadir Cevrim**, who have been amazing friends during my PhD journey in Lyon. **Francesca Andriani**, italian like me, you always made me feel at home, thank you for listening, offering me advices and supporting me throughout this journey.

A huge thank you to **my family**, who always believed in me and in my crazy choices. You always encouraged me and inspired me to follow my dreams. Thank you for making my dreams come true!

Last but not least, this amazing achievement would not have been possible without **my love Gaetano**. I don't have enough words to express my gratitude but thank you for everything, for always be available to travel thousands of km through different countries and making my days brighter. Your love, your support and your constant patience have accompanied me in this amazing journey.





# Table of Contents

Acknowledgments .....	6
List of Figures .....	10
List of Tables.....	12
Résumé en Français.....	14
Summary in English .....	15
List of Abbreviations .....	16
<b>Chapter 1: Introduction</b> .....	<b>18</b>
1. Transcription regulation by enhancers .....	20
1.1 Transcription in eukaryotes.....	20
1.2 The discovery of enhancers .....	21
1.3 General characteristics of enhancers.....	22
1.4 The epigenetic state of enhancers.....	23
1.5 Transcription of enhancers (enhancer RNAs).....	24
1.6 Genome-wide identification of enhancers.....	25
1.7 Characterizing enhancer activity .....	26
1.8 Developmental enhancers.....	27
1.9 Enhancer-promoter interactions .....	29
1.10 Assessing enhancer-promoter interactions: Imaging and 3C based methods .....	30
1.11 Enhancer-promoter interactions and their link to gene expression.....	34
2. Chromatin Organization.....	39
2.1 Chromosome folding.....	39
2.2 Topologically Associating Domains (TADs).....	40
2.3 Polycomb domains.....	42
2.4 Functional role of TADs and their influence on enhancer-promoter communication .....	44
2.5 Molecular mechanisms underlying enhancer-promoter interaction specificity .....	50
3. <i>Drosophila melanogaster</i> early embryogenesis and mesoderm specification .....	57
3.1 Early <i>Drosophila</i> embryogenesis.....	57
3.2 Mesoderm specification.....	58
3.3 <i>twist</i> regulation during early embryogenesis.....	60
4. Concluding remarks .....	62
<b>Aims</b> .....	<b>64</b>

<b>Chapter 2: Results</b> .....	66
1. Introduction.....	68
2. Using the twist E3 enhancer as a model to study enhancer-promoter interactions.....	70
3. The E3 enhancer functions in an orientation-independent manner at its endogenous locus.....	72
4. Generating transgenic fly lines where the E3 enhancer has been inserted in various locations ...	78
5. The deletion of the endogenous E3 enhancer can be rescued by its ectopic re-insertion at given locations.....	85
6. Enhancer-promoter interactions can occur across TAD boundaries and over large distances, even on a different chromosome .....	86
7. Long-range enhancer-promoter interactions are mediated by the presence of the endogenous E3 sequence .....	94
8. The ectopic insertion of the E3 enhancer affects the expression of a large set of genes .....	97
9. Conclusion .....	102
<b>Chapter 3: Discussion</b> .....	106
1. Inverting the orientation of an enhancer at the endogenous locus does not influence the expression of the target gene. ....	108
2. Enhancer and promoter can interact across boundaries and sometimes can rescue the phenotype upon the deletion of the essential regulatory element. ....	109
<b>Chapter 4: Materials &amp; Methods</b> .....	118
1. Material.....	120
1.1 Equipment .....	120
1.2 Plasmids.....	120
1.3 Primers .....	121
1.4 Fly stocks .....	125
2. Methods .....	128
2.1 Generation of transgenic fly lines.....	128
2.2 4C-sequencing .....	135
2.3 3D DNA-FISH .....	138
2.4 RNA-sequencing.....	138
<b>References:</b> .....	141

## List of Figures

Figure 1: Eve gene and its stripe 2 enhancer .....	28
Figure 2: Example of enhancer-promoter chromatin loop.....	30
Figure 3: Outline of the 4C-seq procedure. ....	33
Figure 4: Example of a 4C signal between an enhancer and a promoter in Drosophila. ....	34
Figure 5: Schematic representation of possible mechanism when the gene is OFF and ON. ....	37
Figure 6: Different levels of chromatin organization in the Drosophila genome. ....	42
Figure 7: Regulatory elements in enhancer-promoter communication.....	45
Figure 8: TADs organization. ....	46
Figure 9: Mechanisms that can drive promoter-enhancer specificity.....	55
Figure 10: Diagram summarizing early gene interactions during mesoderm specification. ....	59
Figure 11: Twist expression at three different time points during Drosophila embryogenesis. ....	61
Figure 12: The Twist locus and its three enhancers. ....	62
Figure 13: Characteristics of enhancer E3. ....	71
Figure 14: The deletion of the endogenous E3 enhancer is homozygote lethal.....	72
Figure 15: Schematic example of an enhancer-reporter assay.....	73
Figure 16: Schematic overview of recombinase-mediated cassette exchange. ....	75
Figure 17: The expression of twist is not affected by the orientation of the endogenous E3 enhancer. ..	76
Figure 18: The E3 enhancer probably interacts with the twist promoter irrespective of its orientation. .	77
Figure 19: Overview of the two different experimental strategies used in this study to insert the enhancer E3 at ectopic locations. ....	79
Figure 20: Ectopic insertion sites of the E3 enhancer of twist on chromosome 2R, close to the twist locus. .....	82
Figure 21: Ectopic insertion sites of the E3 enhancer of twist on chromosome 2R, far away from the twist locus.....	83
Figure 22: Ectopic insertion site of the E3 enhancer of twist on chromosome 3L.....	84
Figure 23: The twist promoter doesn't interact with the ectopic E3 enhancer when inserted at locations - 1.6Mb, -181kb, and +39kb. ....	89
Figure 24: The twist promoter interacts with the ectopic E3 enhancer when inserted at locations +7.5kb, +51kb, and on chromosome 3L. ....	91

Figure 25: Distance between twist promoter and the ectopically inserted E3 enhancer at different locations.....	93
Figure 26: The interaction between the twist promoter and the ectopic E3 enhancer is lost upon the deletion of the endogenous E3 enhancer. ....	96
Figure 27: Heatmap of the Spearman correlation coefficient for each RNA-seq library confirms the high quality of the data. ....	98
Figure 28: Effect of the ectopic insertion on the expression of mesodermal genes.....	99
Figure 29: Up-regulated genes are depleted from regions enriched in repressive chromatin marks. ....	101
Figure 31: Map of the pHD-dsRed-2attP vector generated in this study .....	129
Figure 32: Map of the pBS-KS-attB-E3 vector generated in this study.....	130
Figure 33: Map of the pattB-E3 vector generated in this study.....	132
Figure 34: Map of the pBS-KS-attB-E3_MIMIC vector generated in this study .....	133

## List of Tables

Table 1: Three subcategories of enhancers depending on their chromatin features. ....	24
Table 2: Overview of the ectopic insertion sites of the E3 enhancer. ....	81
Table 3: Number of differentially-expressed genes in each condition. ....	98
Table 4: Number of differentially-expressed genes per chromosome in each condition. ....	100
Table 5: Differentially-expressed genes around the insertion site in each condition. ....	100
Table 6: Features of fly lines where the ectopic E3 enhancer is interacting with Twist promoter. ....	113
Table 7: List of equipment. ....	120
Table 8: List of plasmids used in this study. ....	121
Table 9: List of plasmids generated during this study. ....	121
Table 10: List of primers used during this study. ....	124
Table 11: List of MiMIC fly lines using in this study. ....	125
Table 12: List of standard fly lines using in this study. ....	125
Table 13: List of fly lines generated during this study. ....	127
Table 14: List of 4C primers. ....	136



## Résumé en Français

La régulation de l'expression des gènes au cours du développement embryonnaire implique de nombreuses séquences régulatrices (promoteurs, *enhancers*). Les *enhancers* sont de courtes séquences d'ADN situées à une distance variable (proximaux ou distaux) du promoteur de leur gène cible, mais ne régulant pas nécessairement l'expression du gène le plus proche. Pour réguler la transcription, les *enhancers* distaux forment une boucle *enhancer*-promoteur permettant un rapprochement tri-dimensionnel. La proximité 3D entre les promoteurs et les *enhancers* semble être favorisée par la configuration spatiale de la chromatine en domaines d'association topologique (TAD). Cela a conduit à l'idée que les TAD constituent des domaines fonctionnels de base entraînant des interactions *enhancer*-promoteur. Cependant, les mécanismes moléculaires à l'origine de la spécificité d'interaction entre les *enhancers* et les promoteurs à l'intérieur et à l'extérieur des domaines TAD sont encore largement inconnus. Pour identifier ce qui pourrait contrôler la spécificité *enhancer*-promoteur, j'ai systématiquement perturbé le locus de *twist* (le principal régulateur de la fonction du mésoderme) chez les embryons de *Drosophila melanogaster* en utilisant diverses techniques de transgénèse. Tout d'abord, j'ai créé une ligne de mouche où la séquence endogène d'un des *enhancers* de *twist* est présente en orientation inversée. Deuxièmement, j'ai généré six lignées de mouches transgéniques où le même *enhancer* de *twist* est inséré à différents endroits et/ou distances de son gène cible et j'ai vérifié si ces insertions étaient suffisantes pour sauver la létalité de la suppression des *enhancers* endogènes. J'ai ensuite évalué l'impact de ces réarrangements sur l'organisation et la transcription de la chromatine par les méthodes de 4C-seq, 3D DNA FISH et RNA-seq. Les résultats indiquent que chez *Drosophila melanogaster*, les *enhancers* fonctionnent de manière indépendante de l'orientation, même à leur locus endogène. Étonnamment, des interactions *enhancers*-promoteur peuvent parfois se produire au-delà des frontières TAD et même entre un promoteur et un *enhancer* situés sur des chromosomes différents. Enfin, nous avons observé que l'insertion ectopique d'un *enhancer* a un impact énorme sur l'expression des gènes, provoquant un grand nombre de gènes différentiellement exprimés le long du génome. Pris ensemble, ces résultats fournissent une meilleure compréhension de la biologie des *enhancers*, de l'organisation de la chromatine, de la régulation de la transcription, et de la spécificité de l'interaction *enhancer*-promoteur au cours de l'embryogenèse précoce de *Drosophila melanogaster*.

## Summary in English

Embryonic development is controlled by the complex regulation of gene expression, which involves regulatory elements such as promoters and enhancers. Enhancers are short regions of DNA that can be located either proximal or distal to the promoter of their target gene, often skipping closest genes. To regulate gene expression, distal enhancers need to be brought in close three-dimensional proximity to the promoter of their target gene, forming an enhancer-promoter chromatin loop. The 3D proximity between promoters and enhancers seems to be favored by the spatial configuration of chromatin into Topologically-Associating Domains (TADs). This led to the notion that TADs constitute basic functional domains driving enhancer-promoter interactions. However, the molecular mechanisms driving the interaction specificity between enhancers and promoters within and outside TAD domains are still largely unknown. To identify what might control enhancer-promoter specificity, I systematically perturbed the *twist* locus (the master regulator of mesoderm function) in *Drosophila melanogaster* embryos using various transgenesis techniques.

First, I created a fly line where the endogenous sequence of one of the enhancers of *twist* is present in inverted orientation. Secondly, I generated six transgenic fly lines where the same enhancer of *twist* is located at different locations and/or distances from its target gene and checked whether these insertions were sufficient to rescue the lethality of the endogenous enhancer's deletion. I then assessed the impact of these rearrangements on chromatin organization and transcription using 4C-seq, 3D DNA FISH, and RNA-seq respectively. The results indicate that in *Drosophila melanogaster*, enhancers function in an orientation-independent manner, even at their endogenous locus. Surprisingly, enhancer-promoter interactions can sometimes occur across TAD boundaries and even between a promoter and an enhancer located on different chromosomes. Finally, we observed that the ectopic insertion of an enhancer has a huge impact on gene expression, causing a large number of differentially expressed genes along the genome.

Taken together these results provide a better understanding of enhancer biology, chromatin organization, transcription regulation, and enhancer-promoter interaction specificity during early *Drosophila* embryogenesis.



## List of Abbreviations

AEL	After Egg Lay
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
Bcd	Bicoid
bp	Base pair
ChIP-seq	Chromatin immunoprecipitation sequencing
CRM	<i>Cis</i> -regulatory modules
CTCF	CCCTC binding Factor
3C	Chromosome Conformation Capture
3D	Three-dimensional
DEG	Differentially expressed gene
DNase-seq	DNase I hypersensitive sites sequencing
DPE	Downstream Promoter Element
dpp	Decapentaplegic
eRNA	Enhancer RNA
eve	Even-skipped
FISH	Fluorescence In Situ Hybridization
ftz	Fushi tarazu
GFP	Green Fluorescent Protein
Gt	Giant
GRO seq	Global run-on sequencing
Hb	Hunchback
Hbb	$\beta$ -globin
H3K4me1	Histone H3 lysine 4 monomethylation
H3K4me3	Histone H3 lysine 4 tri-methylation
H3K27ac	Histone H3 lysine 27 acetylation
H3K27me3	Histone H3 lysine 27 tri-methylation
H3K36me3	Histone H3 lysine 36 tri-methylation
Ig	Immunoglobulin
Inr	Initiation element

kb	Kilobase
kr	Kruppel
LAD	Lamina Associated Domain
Mb	Megabase
Nc	Nuclear Cycle
oaf	Out at first
PCR	Polymerase Chain Reaction
PcG	Polycomb gene complex
Pol II	Polymerase II
PRC	Polycomb Repressive Complex
PRE	Polycomb Response Elements
RNAPII	RNA Polymerase II
Scr	<i>Sox2</i> control region
Shh	Sonic hedgehog
slp	Sloppy-paired
slh	SLY1 homologus
SNP	Single Nucleotide Polymorphism
TAD	Topologically-Associating Domains
TAL	Transcription Activator-Like effector
TATA	Core promoter element
TF	Transcription Factor
TSS	Transcription Start Site
ZGA	Zygotic Genome Activation

# Chapter 1: Introduction

In this first chapter, I will review current knowledge of three-dimensional (3D) chromosome folding at different scales, highlighting the complex link between genome architecture and transcription. Special attention will be paid to enhancers and to how genome folding into Topologically Associating Domains (TADs) plays a role in enhancer-promoter communication and specificity. Imaging and chromosome conformation capture (3C) methods will be described for their major contributions to the field. I will then introduce *Drosophila melanogaster* as a model system, which has been essential to set up my project. Finally, I will focus on the *twist* gene (master regulator of mesoderm fate) and its enhancers.



## **1. Transcription regulation by enhancers**

### **1.1 Transcription in eukaryotes**

The complexity of a living organism is driven by gene expression regulation, controlling which genes are expressed in which tissue to dictate cell identity. Transcription is the first step of gene expression. Eukaryotic cells have three distinct classes of RNA polymerases, which transcribe different groups of genes. RNA Polymerase I and III mostly transcribe ribosomal RNAs and transfer RNAs respectively, while RNA Polymerase II transcribes protein-coding genes (Cooper, 2000; Wray et al., 2003).

Transcription by RNA polymerase II (RNAPII) can be divided into three main steps: initiation, elongation, and termination. To begin transcribing a gene, RNAPII binds to a specific region lying in the immediate upstream of the transcriptional start site (TSS) of a gene called the promoter. The initiation process of transcription is followed by elongation, where the RNAPII reads the unwound DNA strand and produces a complementary mRNA molecule. Transcription ends when the RNAPII crosses a stop (termination) sequence in the gene, completing the mRNA molecule, which sub-sequentially detaches from the DNA strand (Brown and Clancy, 2008; Orphanides and Reinberg, 2002).

Transcriptional activation is achieved through the binding of sequence-specific transcription factors (TFs) to their binding sites, which are typically clustered in regulatory elements. These regulatory elements influence the transcriptional output positively or negatively, determining when and where genes are turned on and off. Promoters, enhancers, silencers, and insulators are part of the regulatory elements group. As transcription regulation is essential for differential gene expression and thereby development, understanding the mode of action of regulatory elements is crucial.

Genome-wide studies have shown that 97% of the human genome consists of non-protein-coding DNA. In the late 1960s, this part of the genome was considered “junk DNA”. However, the development of new methods has allowed the identification of thousands of regulatory elements located in the non-coding

genomic regions (Alexander et al., 2010). Over 80% of genetic variants and single-nucleotide polymorphisms (SNPs) associated with complex traits and diseases in humans are located in these regulatory elements, suggesting that these variants can affect the control of gene expression. This highlights the importance of the non-coding genome in understanding the regulatory landscape (Alexander et al., 2010; Hrdlickova et al., 2014). Deciphering the *cis*-regulatory code would be essential in order to identify mutations and predict their influence on the organism. This could also help to identify the unknown function of many SNPs in non-coding regions.

## **1.2 The discovery of enhancers**

More than 30 years ago, through the development of genetics and biochemistry methods, it has been shown that DNA sequences located a considerable distance upstream of the TSS of a given gene can play a significant role in transcriptional regulation. Indeed, the deletion of these DNA sequences, named enhancers, had a huge impact on the regulation of a given gene (Atchison, 1988). The first enhancer was identified in 1981 as a 72bp sequence of the SV40 virus genome which could enhance the transcription of a reporter gene by several hundred folds. This enhancement of transcription occurred only when SV40 sequences were present in *cis*, (only on the same DNA molecule) and could occur when the sequences were placed in either orientation or at several thousand base pairs of distance from the gene (Banerji et al., 1981). Studies with the polyoma virus also revealed the existence of a DNA segment essential for viral viability and capable of enhancing the expression of genes in *cis*. Retrovirus were later found to have similar DNA elements in their genome (Atchison, 1988; Tyndall et al., 1981). In 1983, many principles of viral gene regulation mentioned above have also been used to identify enhancers in eukaryotes cells. Indeed, using transient expression assay, the first non-viral enhancer was discovered in the mouse immunoglobulin (Ig) heavy chain gene locus. The enhancer was found to be located within the gene (intron) downstream of the TSS and behaved similarly to the SV40 viral enhancer except for the tissue

specificity. This eukaryotic enhancer is composed of short sequence motifs, which promote the specific activation of the gene in lymphoid cells but not in other cell types (Banerji et al., 1983; Mercola et al., 1983; Serfling et al., 1985).

In 1983, Grosschedl & Birnstiel suggested that

*“The potential for enhancer sequences to provide clues  
to an understanding of differential gene activity in both normal and disease states,  
and their usefulness as experimental tools for examining regulatory systems,  
will ensure their future importance”.*

A statement that turned out to be quite accurate. Since then, many enhancers have been identified and extensively studied (Panigrahi and O’Malley, 2021; Zeitlinger, 2020; Shlyueva et al., 2014).

### **1.3 General characteristics of enhancers**

Enhancers are non-coding sequences in the genome that activate the expression of target genes transcribed by the RNAPII. Enhancers can act independently of their location, orientation, and distance to the target gene (Panigrahi and O’Malley, 2021). Enhancers are bound by cell-type-specific TFs, co-regulators, chromatin modifiers, architectural proteins, and RNAPII, which mediate the activation of the target gene. To allow the assembly of these proteins and enzymes to bind to enhancers sequences, the chromatin has to be accessible (Lewis et al., 2019).

Nucleosomes located in the direct vicinity of active enhancers typically contain histones with characteristic post-translational modifications such as H3K4 methylation and H3K27 acetylation (Bulger and Groudine, 2011; Shlyueva et al., 2014). However, the presence for example of H3K4me1 is not unique to enhancers, as it also coincides with actively transcribed genes (Calo and Wysocka, 2013). As the activity of enhancers is defined by their sequence and can be reproduced outside their endogenous genomic

context (*e.g.* in enhancer-reporter assays), understanding their activation should be an identifiable problem.

#### **1.4 The epigenetic state of enhancers**

Enhancer activity is controlled at three general levels. First, if enhancers lie in a region of compacted chromatin, the region must be converted to a less compacted or open state thanks to the action of pioneering transcription factors. Second, these enhancers must be bound by additional tissue-specific TFs that are critical for the execution of their functions. Third, the epigenetic state of enhancers is fundamental to their function and activity (Small and Arnosti, 2020). Enhancers can be divided into three subcategories depending on their epigenetic status: **(i)** Active enhancers with activation marks such as H3K4me1 and H3K27ac, are bound by the mediator complex, they increase the transcription of target genes and can produce RNA. In this regard, H3K4me1 was the first histone modification linked to distal active regulatory regions through genomic studies in contrast to H3K4me3, which is present at active promoters. This suggests that histone modification patterns can be applied for genome-wide identification of distal enhancers. **(ii)** Primed enhancers, which exist in a primed state before activation, are marked by H3K4me1 and do not transcribe RNA. **(iii)** Poised enhancers are similar to primed enhancers, however, they are also marked by H3K27me3 and bound by PRC2 (Polycomb Repressive Complex 2) (Table 1). These elements are located near early developmental genes and were first described in human and mouse embryonic stem cells (Rada-Iglesias et al., 2011; Bonn et al., 2012; Calo and Wysocka, 2013; Sharifi-Zarchi et al., 2017). Taken together, these pieces of evidence show that histone modification features can be used to identify unknown enhancers.



Active Enhancers	Primed Enhancers	Poised Enhancers
H3K4me1	H3K4me1	H3K27me3
H3K27ac		

**Table 1: Three subcategories of enhancers depending on their chromatin features.** Active enhancers are bound by H3K4me1 and H3K27ac. Primed enhancers are marked by H3K4me1 prior to activation. Poised Enhancers are bound by the repressive chromatin mark H3K27me3.

**1.5 Transcription of enhancers (enhancer RNAs)**

Additional factors important for the identification of active enhancers are the binding of RNAPII and the production of non-coding enhancer RNAs (eRNAs). The discovery that enhancers are transcribed to produce non-coding RNAs has added another layer of complexity to the mechanisms of their activity and has generated an intense area of investigation (Lewis et al., 2019). Since the discovery of eRNAs, transcription is now considered a general feature of active enhancers. Enhancer RNAs are short RNAs that are transcribed from open chromatin regions that result from transcription factor occupancy of enhancer regions (Henriques et al., 2018; Kim et al., 2010). Although the presence of eRNAs has proven to be a strong indicator of activity, it has been challenging to define clear roles for eRNAs in enhancer function. eRNAs often initiate at the edges of the nucleosome-depleted regions, with many enhancers being transcribed in both directions. However, the instability of eRNAs makes them difficult to detect using RNA-sequencing. Therefore, other methods such as global run-on sequencing (GRO-seq) and Start-sequencing isolate nascent RNAs during the act of transcription, while they are still engaged with RNAPII and are protected from degradation. These assays also provide a sensitive approach for the identification of enhancers (Core et al., 2014; Henriques et al., 2018).

## 1.6 Genome-wide identification of enhancers

Different genome-wide methods can be used to detect enhancers. For example, chromatin immunoprecipitation followed by sequencing (ChIP-seq) is widely used to identify *in vivo* protein-DNA binding events. Since many TFs bind to enhancer regions, enhancers can be identified using TF ChIP-seq. Although this method can be useful to identify putative enhancer regions, it also tends to overestimate the number of true enhancers. Indeed, TFs can also bind to other *cis*-regulatory regions and to non-functional elements in the DNA challenging the identification of true enhancers (Abdulghani et al., 2019). As enhancers require protein-binding to exert their regulatory functions, the chromatin has to be accessible. Therefore, chromatin accessibility is a defining feature of active enhancers and can be measured by using various techniques such as DNase I hypersensitivity (DNase-seq) or transposase-assisted measurement of chromatin accessibility assay (ATAC-seq). However, accessibility methods can indicate a very large region containing multiple regulatory elements and are not always indicative of an active enhancer. Indeed, these approaches have proven that no single genomic feature is highly predictive of an active enhancer. However, CRISPR methods in combination with single-cell RNA sequencing have confirmed that H3K27ac, p300, and RNAPII occupancy truly mark active enhancer sites (Field and Adelman, 2020).

Moreover, some enhancers are also bound by CTCF and cohesin, which have been proposed to mitigate communication between an enhancer and its target promoter by forming chromatin loops (Phillips and Corces, 2009). Although several specific enhancers could activate promoters without forming chromatin looping, it is widely known that the majority of enhancers come into spatial interactions with promoters of their target genes. Indeed, additional technologies such as chromosome conformation capture (3C) based methods, which include 4C/Hi-C/Promoter-Capture, aim to study the chromatin organization either at specific loci or on the entire genome. In particular, these technologies have been widely used to identify putative enhancers (Golov et al., 2020).

However, these assays individually may not provide a strong prediction of enhancer elements. A possible solution is to combine these different types of experiments with several computational studies which gather published datasets from multiple cell types and fit them into statistical models to predict active enhancers and their target promoters (Abdulghani et al., 2019; Lewis et al., 2019; Field and Adelman, 2020).

### **1.7 Characterizing enhancer activity**

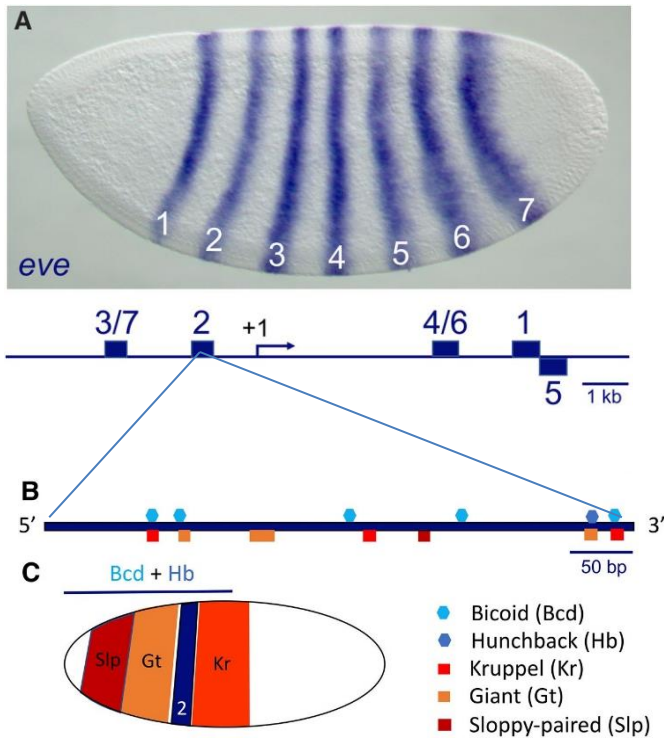
There is no unique assay to definitively characterize enhancer elements and identify their target genes. The methods mentioned above are important tools to predict enhancer location, however, the most common method for evaluating the potential of a given sequence to act as an enhancer is enhancer-reporter assays. The construct used in these assays includes a reporter gene driven by a minimal promoter, which alone is insufficient to drive the expression, and a putative enhancer cloned upstream of the minimal promoter. However, reporter assays cannot link enhancers to their target genes and do not reflect the native chromatin architecture of their endogenous locus. To understand the biological role of enhancers in their endogenous chromatin environment, experiments like enhancer traps are heavily used. This technique allows the random insertion of a reporter gene construct into the genome by a transposase. Inserts within range of an enhancer will be activated in a manner that reflects the spatial and temporal activity of the enhancer, which is evaluated for example by staining embryos at specific time points. While enhancer traps do not directly identify enhancers or their target genes, they emphasize the importance of the local chromatin context in enhancer function. For this, the ability to alter putative enhancers in their endogenous genomic environment is required. CRISPR-Cas9 mutagenesis directed by guide RNAs has become a flexible and reliable tool for performing such perturbations (Field and Adelman, 2020). Overall these assays together can provide insight into the enhancer function of a genomic locus.

## 1.8 Developmental enhancers

Embryonic development is controlled by the complex regulation of gene expression, directing a cell to establish a particular fate. This control involves *cis*-regulatory elements, such as enhancers. Developmental enhancers are typically 200bp – 1kb in length and can switch on/off patterns of gene expression, in specific cell types and at particular stages during development (Levine, 2010). These types of enhancers can contain many binding sites for TFs, which determine when and where the enhancer activates transcription. Developmental enhancers are thus essential to achieve the precise regulation of spatiotemporal gene expression (Small and Arnosti, 2020). As the function of enhancers during development is stage- and tissue-specific, identifying regulatory elements in embryos using genomics methods can be difficult due to the limited amount of material available. Model organisms such as *Drosophila*, where large amounts of cells can be obtained, are therefore an excellent model to identify developmental enhancers (Koenecke et al., 2016).

One of the most well-known and best-characterized developmental enhancers in *Drosophila* is the *even-skipped (eve)* stripe 2 enhancer. The *eve* gene encodes a homeodomain protein expressed in a series of seven stripes that controls the segmentation of *Drosophila* embryos. *Eve* is regulated by multiple enhancers, each responsible for a specific stripe of the complete expression pattern (Figure 1A) (Farley and Levine, 2012). The *eve* stripe 2 enhancer contains binding sites for transcription factors such as Bicoid (Bcd) and Hunchback (Hb) that activate transcription, but also Sloppy-paired (Slp), Kruppel (Kr), and Giant (Gt) that function as repressors. Each transcription factor binds its corresponding DNA motif and independently directs the expression of one or two stripes (Figure 1B). Bcd and Hb are expressed throughout the anterior half of the embryo, whereas the repressors are expressed in specific domains along the anterior-posterior axis of the embryo (Figure 1C). Therefore, loss of any of the binding sites can affect the *eve* expression pattern in *Drosophila* embryos. For example, the loss of Gt binding sites causes an anterior expansion of the normal pattern (Farley and Levine, 2012; Small and Arnosti, 2020). Overall,

gene expression patterns are regulated by multiple enhancers which are responsible for the development of the embryo.

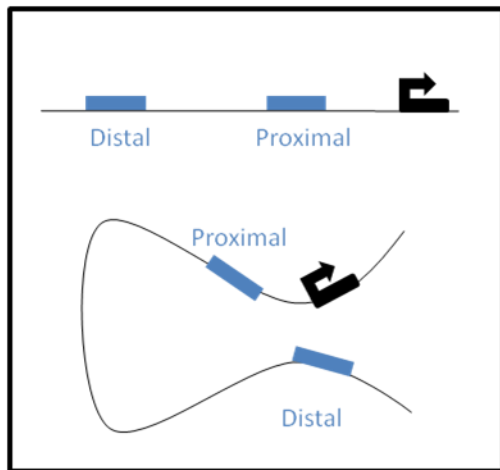


**Figure 1: Eve gene and its stripe 2 enhancer.** **A.** *In situ* hybridization of *Drosophila* embryo to detect mRNAs. Each stripe pattern is regulated by a specific enhancer. The blue rectangles indicate the position of enhancers in the genome, TSS is indicated by +1 arrow. **B.** The stripe 2 enhancer shows binding sites for 5 different TFs, which are Bcd, Hb, Slp, Kr and Gt. **C.** Schematic representation of a *Drosophila* embryo showing the expression pattern of each TF. Adapted from (Small and Arnosti, 2020).

## 1.9 Enhancer-promoter interactions

Because enhancers and promoters can be separated by many kilobases along the genomic sequence, a major challenge is to determine which regulatory elements control which genes. This is difficult because regulatory elements can control multiple genes and can be scattered over a large distance, often skipping intervening genes (van Arensbergen et al., 2014). Four different models have been proposed to explain how enhancers interact with their target promoters. **(i)** The first model is called the 'tracking model'. A protein binds to the enhancer and tracks the enhancer along the chromatin fiber in the direction of the promoter, inducing transcription. **(ii)** The second model is the 'linking model'. The loaded protein on the enhancer sequence drives the polymerization of proteins towards the promoter. **(iii)** The third model is the 'relocation model'. A given gene relocates to nuclear compartments mediating enhancer-promoter interactions **(iiii)** The fourth model is the 'looping model'. A distal enhancer creates a chromatin loop with its promoter (Kolovos et al., 2012). Over the years, the looping model has become the most prominent and vastly accepted model.

Even in the compact *Drosophila* genome, around 30% of enhancers work over distances of 70-100kb (Furlong & Levine, 2018). Indeed, long-range interactions are highly common, and each enhancer interacts with multiple enhancers and promoters with a similar expression, indicating a role in their co-regulation. When enhancers are located at large distances, up to several hundred kilobases or even a megabase, in order to function, they need to be brought close to their target promoter(s) to activate transcription via the release of RNAPII. This is achieved via the formation of an enhancer-promoter chromatin loop (Figure 2) (Moretti et al., 2019; Schoenfelder and Fraser, 2019; Shlyueva et al., 2014).



**Figure 2: Example of enhancer-promoter chromatin loop.** Enhancers can be located close or far from their target gene (Top). To function, distal enhancers need to be brought in close proximity to the promoter of their target gene forming a chromatin loop (bottom).

An example of long-range interactions, where an enhancer skips a nearby gene to activate a more distal gene is seen in the *Drosophila* Antennapedia complex. The T1 enhancer located at the 3' of the *fushi tarazu* (*ftz*) gene can bypass the gene in order to activate the more distal *Sex combs reduced* (*Scr*) gene (Levine, 2010).

However, how enhancers find and interact with a distal promoter to trigger transcription and what stabilizes these interactions remains still under deep investigation. Recently, understand the mechanisms behind long-range interactions has become more feasible thanks to new developments in genomics (Chromosome Conformation Capture technology) and high-resolution imaging.

### 1.10 Assessing enhancer-promoter interactions: Imaging and 3C based methods

Understanding the 3D genome organization is important in order to determine the chromosomal activities within the cell. Hence, a fundamental question in genome biology is how the spatial conformation of the chromosome in the nucleus affects several genetic and biological functions such as gene regulation, DNA

repair, and DNA replication. Studies of enhancer-promoter interactions and chromosome organization mostly rely on two types of methods: imaging techniques, such as DNA fluorescent *in situ* hybridization (FISH), and next-generation sequencing techniques such as chromosome conformation-based methods (Oluwadare et al., 2019).

#### *(A) Imaging Nuclear Organization – DNA FISH*

Imaging has been a dominant method to define genome folding. Fluorescence microscopy allowed the discovery of many principles of chromosome organization, including the formation of chromatin loops and chromatin domains. Furthermore, imaging presents certain advantages compared to genomics-based methods. It gives the ability to probe chromatin folding within tissues, keeping the information of the spatial organization of cells within their endogenous context. For many years, the structure of the genome has been studied based on microscopy methods. Chromosome organization can be probed by DNA FISH, a method that uses fluorescent probes hybridizing on specific genomic loci in single-cell nuclei and detected by confocal microscopy. In particular, FISH allows the visualization of genomic loci within the nuclei and is efficient to assess the variability of chromatin folding within a population of cells. The findings have helped the understanding of the genome architecture and behavior in the nucleus of the cell. However, due to the limit of resolution in conventional microscopy, FISH has been used to probe the relative positioning of genomic loci by measuring distances rather than directly visualizing the structure of the labeled region. Recently, the development of super-resolution microscopy approaches has helped to overcome these issues. However, to provide deep information on the chromatin organization of specific genomic loci, FISH remains a technique that can be used in combination with chromatin conformation capture (3C) based methods (Oluwadare et al., 2019).



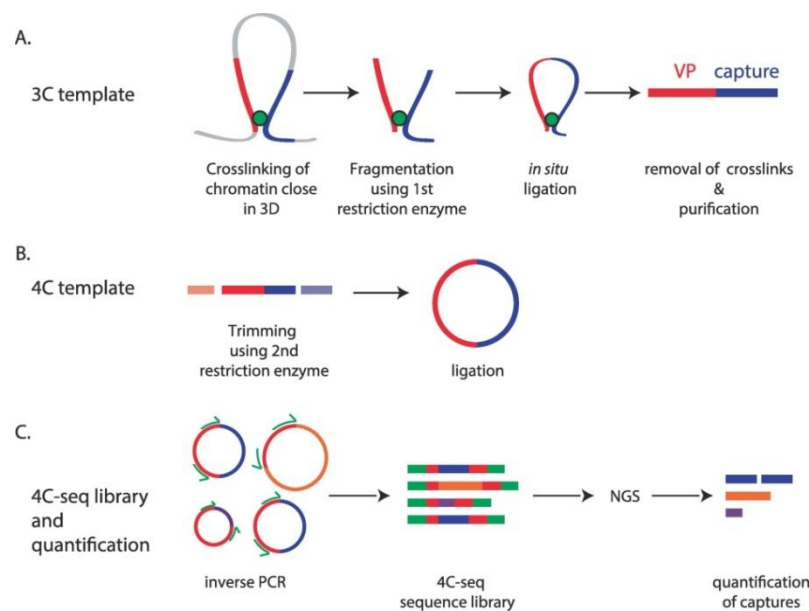
### *(B) Chromosome Conformation Capture*

Studies of chromatin organization within the nucleus have been facilitated by the 3C techniques and their extensions. The 3C-based methods are based on measuring contact frequencies between pairs of chromatin fragments, which are located in close spatial proximity. Nuclei are crosslinked using formaldehyde before digestion of chromatin by a restriction enzyme. Chromatin fragments that are in close spatial proximity are then ligated to each other using a DNA ligase and then detected by PCR or next-generation sequencing (Dekker et al., 2002). Hence, genomic loci that are frequently in contact will likely be more often detected in a population of cells than those that rarely interact. One of the 3C-derived methods, called 4C-seq, allows the detection of interactions between a locus of interest called the viewpoint and multiple other loci (Figure 3-4). It is usually the method of choice to study enhancer-promoter interactions, given its very good but more localized resolution compared to genome-wide methods such as Hi-C. However, 3C-based methods rely on the use of crosslinking and ligation, which could potentially cause experimental bias, questioning whether interactions detected by the 4C method actually exist in living cells. Therefore, an alternative assay to study chromosomal contacts without biochemical manipulation of cells has recently been developed (Redolfi et al., 2019). This method is named DamC, and associates DNA methylation with the identification of chromosomal contacts by next-generation sequencing and uses biophysical modelling for methylation kinetics. DamC provided novel insights into chromosomal structure, and confirmed that the same interactions can be detected by the 4C method. This indicates that crosslinking and ligation are not an issue for the detection of chromatin contacts (Redolfi et al., 2019).

Overall, chromosome conformation capture techniques have provided deep knowledge in genome organization. However, these methods generate cell population interaction profiles, which make difficult the interpretation of single chromosomal structures. Thus, many studies have combined 3C based

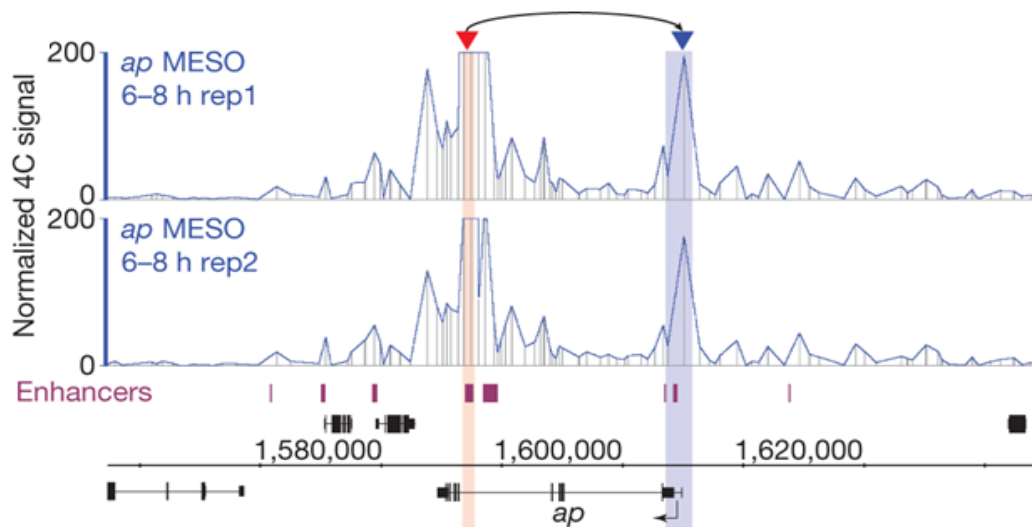
methods with DNA FISH assays, which can be used to measure distance distributions from individual cells between loci of interest (Giorgetti and Heard, 2016). With the development of super-resolution microscopy and new generations of probes, DNA FISH and other imaging methods have become very powerful tools to characterize chromatin conformation in individual cells (Fudenberg and Imakaev, 2017). Furthermore, thanks to the development of single-cell Hi-C, chromatin interactions in individual cells have been identified, which allow a deeper insight into chromatin organization in different cell types (Nagano et al., 2013).

Together these techniques have contributed to the field of enhancer biology but also have confirmed the importance of chromatin organization in gene regulation.



**Figure 3: Outline of the 4C-seq procedure.** **A.** Genomic regions that are spatially proximal in the cell nucleus (red and blue) are crosslinked (green). Chromatin is fragmented using a primary restriction enzyme, fragments in close proximity are ligated *in situ*, after which the crosslinks are reversed and the resulting molecules (the 3C template) are purified. **B.** In 4C-seq the 3C template is trimmed using a secondary restriction enzyme followed by circularization by a second ligation. **C.** To identify and quantify

fragments that are ligated (blue, orange, and purple) to the genomic region of interest (the viewpoint, red), an inverse PCR is performed, and the amplicons are analyzed using next-generation sequencing (Krijger et al., 2020a).



**Figure 4: Example of a 4C signal between an enhancer and a promoter in *Drosophila*.** 4C interaction map (viewpoint, red arrowhead) at the *ap* locus. The expected interaction with the promoter (blue arrowed) of *ap* is observed. Known enhancers are indicated. Adapted from (Ghavi-Helm et al., 2014).

### 1.11 Enhancer-promoter interactions and their link to gene expression

While multiple studies documented differences in enhancer-promoter interactions between different cell types and tissues, it is still unknown whether gene expression and chromatin state drive these interactions or whether changes in the proximity between an enhancer and a promoter facilitate tissue-specific gene expression. Many studies have observed that enhancer-promoter interactions are formed simultaneously with gene expression, without being able to reveal whether enhancer-promoter interactions are the cause

or the consequence of gene activation. Recently it has been shown that forcing the chromatin loop between an enhancer and a promoter can lead to strong transcriptional activation. For example, experimentally induced chromatin contact between the mouse  $\beta$ -globin (*Hbb*) promoter and its enhancer causes the activation of the *Hbb* gene, showing that enhancer-promoter contacts can activate transcription (Deng et al., 2012).

In particular, tissue-specific genes such as the  $\beta$ -globin gene create chromatin interactions depending on the status of their expression. For example, in erythroid cells, the  $\beta$ -globin gene interacts with active regions, whereas in fetal brains, when the gene is not active, interactions occur in inactive regions. On the contrary, the housekeeping gene *Rad23a* which is located in an active chromosomal region makes contacts with active regions in *cis* and *trans*, and the profile of the contacts is conserved between different tissues (Wit and Laat, 2012).

Moreover, a recent study used as a model system the dorso-ventral patterning of the *Drosophila melanogaster* embryo to demonstrate that there is independence between chromatin organization and dorso-ventral gene expression (Ing-Simmons et al., 2021). In detail, they identified tissue-specific enhancers linked to expression patterns in *Drosophila* embryos using single-cell RNA sequencing.

Notably, although each tissue showed specific chromatin states and gene expression, chromatin organization was widely preserved across tissues. Consequentially, this indicates that tissue-specific chromatin conformation is not necessary for tissue-specific gene expression but rather acts as a scaffold mediating gene expression when enhancers become active (Ing-Simmons et al., 2021). This result has been also supported by a recent study that developed a new imaging technique called Hi-M, which enables the detection of chromatin organization and transcription states in *Drosophila* embryos. More specifically, this technology visualizes where and when enhancer-promoter interactions occur and study their effects on transcriptional status (Espinola et al., 2021). Thanks to this new technique, *cis*-regulatory modules (CRMs), which include enhancers and target promoters interactions, have been identified in

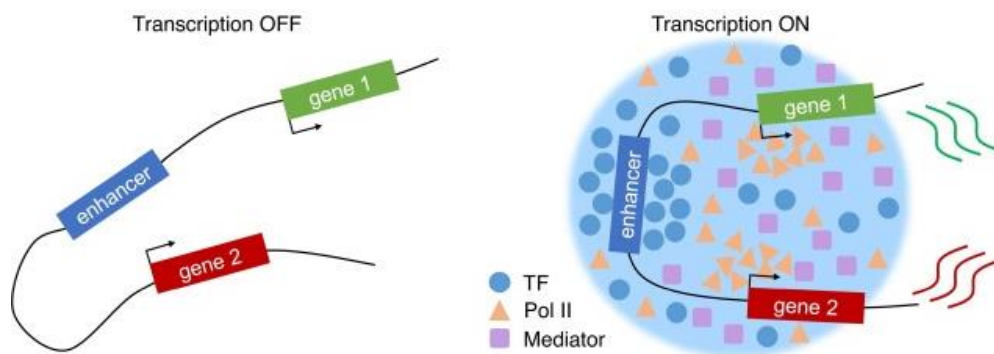
different cell types at early stages of embryogenesis in *Drosophila*, particularly at nc11, and remained almost unaffected at least until nc14, where the formation of chromatin domains and transcription activation occurs. Interestingly, one of the main findings is that these CRMs with multiple enhancers' clusters create *cis*-regulatory hubs. These hubs are mainly invariant during cell fate specification and gene activation during *Drosophila* embryogenesis. This indicates that it might be independence between CRMs and gene expression. For example, enhancer-promoter interactions can be detected in tissues where target genes are not expressed. From a developmental point of view, the formation of these interactions in cells where genes need to be repressed can be seen as a problem. Indeed, once a loop is formed, transcriptional activation could occur in cells where the gene is supposed to be inactive. In other cases, even with the present of enhancer-promoter interactions, transcriptional activation does not occur and enhancers can act as silencers in different cell types. However, this does not exclude other silencing mechanisms that can play a role (Espinola et al., 2021).

At other loci, enhancer-promoter interactions seem both dynamic and cell-type-specific and are closely associated with changes in gene expression (Noordermeer and Laat, 2008). In both situations, the interactions are at least partially formed prior to transcription. This indicates that enhancer-promoter interactions are essential, but not sufficient to activate a gene.

While most studies indicate that enhancer-promoter proximity is essential for gene expression, a recent result suggests that enhancers could also potentially control gene expression without direct contact with the promoter. For example, sonic hedgehog (*Shh*) expression in the developing mouse brain is regulated by several *Shh* enhancers. Surprisingly, when the gene *Shh* is expressed, no contacts are detected between the *Shh* promoter and these enhancers (Benabdallah et al., 2019). This suggests that an enhancer may not necessarily need to be in continuous physical proximity with a promoter to influence transcription (Schoenfelder and Fraser, 2019; Panigrahi and O'Malley, 2021). The common model is that distal

enhancers loop to target gene promoters where they stimulate transcription by providing an increased local concentration of transcription factors.

Indeed physical distance between an enhancer and its target promoter becomes larger, not smaller, upon transcription activation. This suggests that the activity of the gene can influence the distance between enhancer and promoter. A model that could explain this phenomenon could be that, upon the activation of a gene, a large transcription hub starts to be formed, which includes RNAPII, TFs, and Mediator. This transcriptional machinery causes an increase in the distance between enhancers and promoters but at the same time can facilitate the interaction between enhancer and promoter (Figure 5)(Lim and Levine, 2021).



**Figure 5: Schematic representation of possible mechanism when the gene is OFF and ON.** When transcription is ON, transcription factors Polymerase II and Mediators increase the distance between the enhancer and its target genes creating a machinery hub (blue circle). Adapted from (Lim and Levine, 2021).

Nevertheless, this model does not function to all cases. Sometimes, distal enhancers prefer to interact with the intronic sites their target genes (e.g. at) rather than with promoter regions (Stadhouders et al., 2012). The mechanism, duration, and timing of such contacts with respect to gene activity are still under debate.

As it is mentioned before, the traditional model for transcription involves the active form of the polymerase tracking along the DNA template. However, sometimes the polymerase is concentrated with other factors forming a machinery factory. The term 'transcription factory' is used to describe a locus containing at least the RNA polymerase and protein-binding sites clustering together into 'hubs'. This transcription factory can play an important role in maintaining chromatin loops, for example when 3C based methods detect *trans*-interaction (between chromosomes), it can be a result of transcription factories which help to tie the interaction through the recruitment of different transcription factors. However, this does not exclude that other loops will be stabilized by other molecular components outside factories (Papantonis and Cook, 2013).

Live-cell super-resolution imaging of labelled RNAPII has revealed that clustering of the polymerase near the promoter is extremely transient, suggesting that enhancer-promoter interactions are quite dynamic. (Cho et al., 2016). In support of this study, it has recently been shown that active gene and enhancer regions, containing decondensed chromatin, are often more mobile and dynamic than inactive regions. This suggests that upon chromatin opening, the enhancer and promoter regions increase their mobility rate allowing more frequent contacts between these regulatory elements (Gu et al., 2018). Another live-cell imaging study of *Drosophila* showed that at the *eve* locus, transient enhancer-promoter proximity is correlated with gene activity. The lifetime of enhancer-promoter colocalization increases when transcription is activated, supporting the previously mentioned model where the high local concentration of TFs at promoters and enhancer that accompanies bursts of transcription, facilitates transient enhancer-promoter communication. However, as these studies have evaluated enhancer-promoter interactions using cell imaging methods, and none at endogenous loci, further development of technologies is needed to understand the functional role of these regulatory elements in gene regulation.

## 2. Chromatin Organization

### 2.1 Chromosome folding

In the nucleus of eukaryotic cells, chromosomes adopt a highly organized structure and occupy distinct territories. Chromosomes fold in the three-dimensional space of a cell nucleus in a non-random fashion. Different chromosomes occupy distinct regions of the nucleus referred to as chromosome territories. At the periphery of the nucleus, chromatin is associated with the nuclear lamina and appears very compact and generally transcriptionally inactive, while close to the center of the nucleus chromatin it is less compact and enriched in transcriptionally active regions (Dekker et al., 2002; Moretti et al., 2019; Rowley and Corces, 2018). *Drosophila* chromosomes present additional characteristics. First, they adopt a polarized orientation, known as the Rabl configuration, with centromeres and telomeres clustering separately at two opposite nuclear poles (**Figure a**) (Moretti et al., 2019; Padeken et al., 2013; Pickersgill et al., 2006). Second, in *Drosophila*, homologous chromosomes are paired during the cell cycle and interact with each other through *trans*-interactions. Consequently, gene expression can be affected by a mechanism called transvection, where *cis*-regulatory elements, which include enhancers and/or promoters, regulate the expression of genes located on the homologous chromosome (Lim et al., 2018). At the sub-megabase scale, chromosomes segregate in two compartments called A and B compartments. The A compartment contains highly transcriptionally active genes and active chromatin marks whereas the B compartment is composed of transcriptionally repressed chromatin regions that are closed and tightly packed (**Figure b**). This large-scale organization is dynamic during cell differentiation, raising important questions about the basis and functional role of genome folding heterogeneity (Moretti et al., 2019).



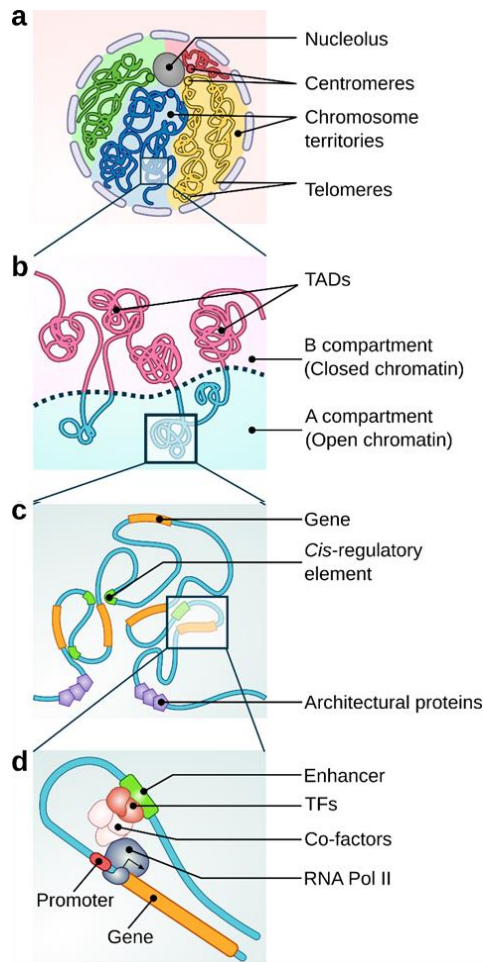
## 2.2 Topologically Associating Domains (TADs)

Each compartment A and B contains multiple chromatin loops, which form self-associating domains, called Topologically Associating Domains (TADs) (Figure 6b-c-d) (Moretti et al., 2019; Szabo et al., 2019). TADs are genomic regions where physical chromatin interactions occur frequently within a domain, but more infrequently across domain boundaries (Figure 6b-c). Within TADs, chromatin loops between regulatory elements such as enhancers and promoters, are implicated in the fine regulation of gene expression. The presence of these domains has been described in many species, indicating that they may represent robust functional units of gene regulation, affecting enhancer-promoter interactions.

In mammalian genomes, TAD domains are several hundred kb in size, up to 1-2Mb, whereas they are smaller in bacteria (from 30 to 420kb) and flies (about 60kb), (Sexton and Cavalli, 2015; Szabo et al., 2019). However, the detection of these domains is highly dependent on the resolution of the Hi-C method that generates contact maps between all parts of the genome and on the computational method used to define TAD calls. More specifically, the threshold used in the TAD algorithms can influence the discovery of TAD domains, generating a large discrepancy between studies. In addition, TADs strongly correlate with epigenetic domains. While chromatin state is consistent within the same TAD, the segregation of active and repressive marks is not always perfect and TADs can display heterogeneous chromatin states (Moretti et al., 2019). Therefore, it is possible to classify TADs based on their chromatin states. Active TADs are enriched in transcriptionally active genes displaying chromatin marks such as H3K4me3 and H3K36me3. Inactive TADs are more frequently found associated with Polycomb repressed loci enriched in H3K27me3 and with the nuclear lamina displaying chromatin marks such as H3K9me2 (Sexton et al., 2012).

The large majority of TAD boundaries are present in transcribed regions with active chromatin marks and are mainly enriched for insulators or architecture proteins binding, such as the CCCTC-binding factor (CTCF) and the cohesin complex (Figure 6c) ((Moretti et al., 2019; Schoenfelder and Fraser, 2019). In vertebrates, CTCF is the predominant insulator protein and the orientation of its binding site with respect

to another CTCF site is involved in the formation of chromatin loops. When one of these sites is mutated or inverted, the loop will form with the next convergent site, skipping divergent sites of CTCF. This observation has led to the description of a “pairing rules” model (Cavalheiro et al., 2021). In *Drosophila*, many insulator proteins have been discovered, including CTCF, BEAF-32, CP190, Zw5, Pita, ZIPIC, Ibf1/2, Su(Hw), and Opbp. Although the insulator “pairing rules” in flies have yet to be deciphered, there are examples where, some insulators act to block enhancer cross-talk, while others help to bring enhancers close to their target promoters (Cavalheiro et al., 2021). Moreover, in contrast to mammals, in *Drosophila*, there is little enrichment of CTCF at TAD boundaries. The reason why fly CTCF is not a major insulator element is still unknown (Szabo et al., 2019). However, CTCF in *Drosophila* plays an important role in *Hox* genes regulation. Indeed, flies missing zygotic CTCF do not survive as young adults due to misexpression of *Hox* genes in incorrect body segments, which cause homeotic defects (Gambetta and Furlong, 2018). Genome-wide studies have shown that TAD boundaries are more evolutionary conserved across cell types compared to the sequences within the TAD domain. As TAD boundaries are enriched for housekeeping genes and transcription start sites, mutations located at TAD boundaries most likely also affect the genes and enhancers that are present at these boundaries, causing both rare and common diseases. Therefore, a deeper insight into TAD formation will be critical to further understanding the role of TADs in gene expression regulation, and decipher their importance in various pathologies (McArthur and Capra, 2021).



**Figure 6: Different levels of chromatin organization in the *Drosophila* genome.**

**A.** Chromosomes are compacted within the nuclei, the centromeres of the chromosomes cluster together to one pole of the nucleus close to the nucleolus, while the telomeres are found in the opposite pole.

**B.** Chromosomes based on their chromatin state can be organized in A (blue) and the B (pink) compartments, which each of them contains TADs domains. **C.** TADs bring into close proximity genes and their cis-regulatory elements. TADs are delimited by the presence of architectural protein binding sites at their borders. **D.** Within TADs, enhancers are brought into close proximity to the promoter of the genes they regulate through loops that can be mediated by insulator pairing, TFs, co-factors or RNA Pol II pausing (Moretti et al., 2019).

### 2.3 Polycomb domains

Polycomb Repressive Complexes are epigenetic regulators important for gene expression. PRC1 mono-ubiquitylates the histone H2A tail on residue K119, and PRC2 methylates the histone H3 tail on residue K27, producing H2AK119ub1 and H2K27me3, respectively. Both histone marks can influence chromatin compaction and gene silencing. PRCs are proteins that can help the formation of specific types of repressive TADs named Polycomb domains. In *Drosophila*, Polycomb group proteins are recruited on chromatin by Polycomb Response Elements (PRE), which act as nucleation sites for the formation of repressive Polycomb domains marked by H3K27me3. Hi-C data showed that in *Drosophila*, repressive

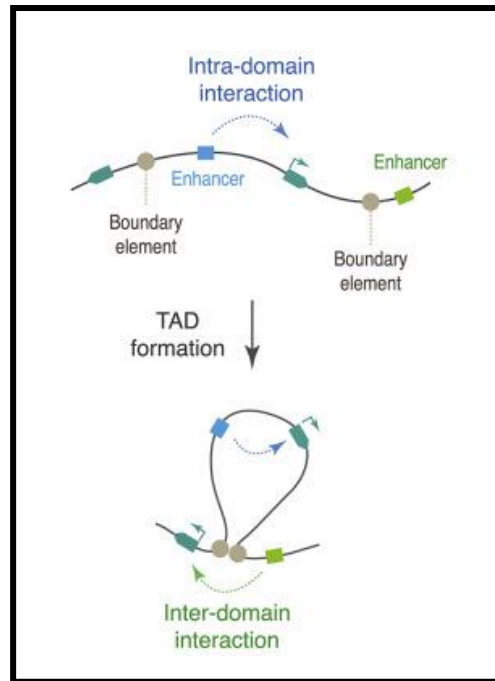
loops within Polycomb domains are formed after the midblastula transition between PREs and the action of the transcription factor GAGA and Polycomb proteins. Perturbation of PRE function by CRISPR-Cas9 editing affects Polycomb domain formation and destabilizes Polycomb-mediated silencing (Ogiyama et al., 2018). This points out that Polycomb proteins have roles as organizers of chromatin architecture, revealing a new layer of Polycomb-mediated regulatory mechanisms. In more detail, they have been shown to drive interactions between enhancers and promoters. To investigate the role of Polycomb in *cis*-regulatory interactions, knockout experiments of two components of the PRC2 complex (EED and SUZ12) disrupted the contacts between genes and poised enhancers, highlighting that PRC2 itself and its associated mark H3K27me3 are important for these interactions (Caglio et al., 2017). Moreover, at the TAD scale, super-resolution imaging of Polycomb domains in *Drosophila* indicated that they are more condensed than active or inactive domains. This condensation was shown to rely on PRC1 both in *Drosophila* and mammals. Interestingly, the loss of chromatin compaction upon the removal of PRC1 components can be observed independently of transcriptional activation, which confirms an important role for these proteins in establishing a 3D chromatin environment (Cheutin and Cavalli, 2019). In addition, Polycomb domains have been shown to undergo long-range contact networks. Indeed, in *Drosophila*, FISH labeling of the two *Hox* gene clusters, which are separated by 10Mb showed that they can contact each other within the nucleus, resulting in the formation of one large Polycomb body in which *Hox* genes are repressed. Perturbation of this interaction by mutating one of the two loci results in the silencing of the other one. This highlights the function of Polycomb long-range contacts in stabilizing transcriptional repression (Bantignies et al., 2011). Furthermore, binding sites of PcG proteins undergo specific looping involving gene promoters and *cis*-regulatory elements that contribute to domain compaction and efficient gene silencing. Therefore, this functional organization of Polycomb proteins plays an important role in gene regulation.

## 2.4 Functional role of TADs and their influence on enhancer-promoter communication

In *Drosophila melanogaster*, the stages of development are characterized by 13 rapid nuclear divisions until the cellularized embryo undergoes zygotic genome activation (ZGA) at nuclear cycle 14 (Hamm and Harrison, 2018). To examine how chromatin is organized when the zygotic genome is not yet transcribed, *in situ* Hi-C maps have been performed in embryos at around ZGA time. Before ZGA, the genome does not have a defined structure. Only upon the recruitment of RNAPII, TADs start to be established throughout the genome and co-localization of housekeeping genes at TAD boundaries starts to occur. However, upon inhibition of RNAPII activity by pharmacological treatment, TAD formation is not affected, but contacts properties of TADs and housekeeping genes enriched at TAD boundaries are perturbed. As a result, although TAD formation is independent of transcription, transcription is essential to control correct chromatin organization (Hug et al., 2017).

In other cases, TADs can be present before gene expression. Indeed, the  $\alpha$ -globin locus in mammals creates a self-interacting domain before the  $\alpha$ -globin gene is actually transcribed. This indicates that transcription is not a prerequisite to the formation of TADs (Brown et al., 2018). Similarly, other studies have shown that, during the cell cycle, gene activation occurs before the establishment of topological domains (Zhang et al., 2019).

To better understand the role of TAD domain and its relation to gene regulation, two probable functions of TAD domains have been suggested. (A) TAD boundaries act as a boundary in the classic insulator sense, whereby they block inter-TAD enhancer-promoter communication; (B) TADs facilitate intra-TAD communication, which can increase the frequency and stabilize enhancer-promoter interactions (Figure 7) (Cavalheiro et al., 2021; Yokoshi et al., 2020).

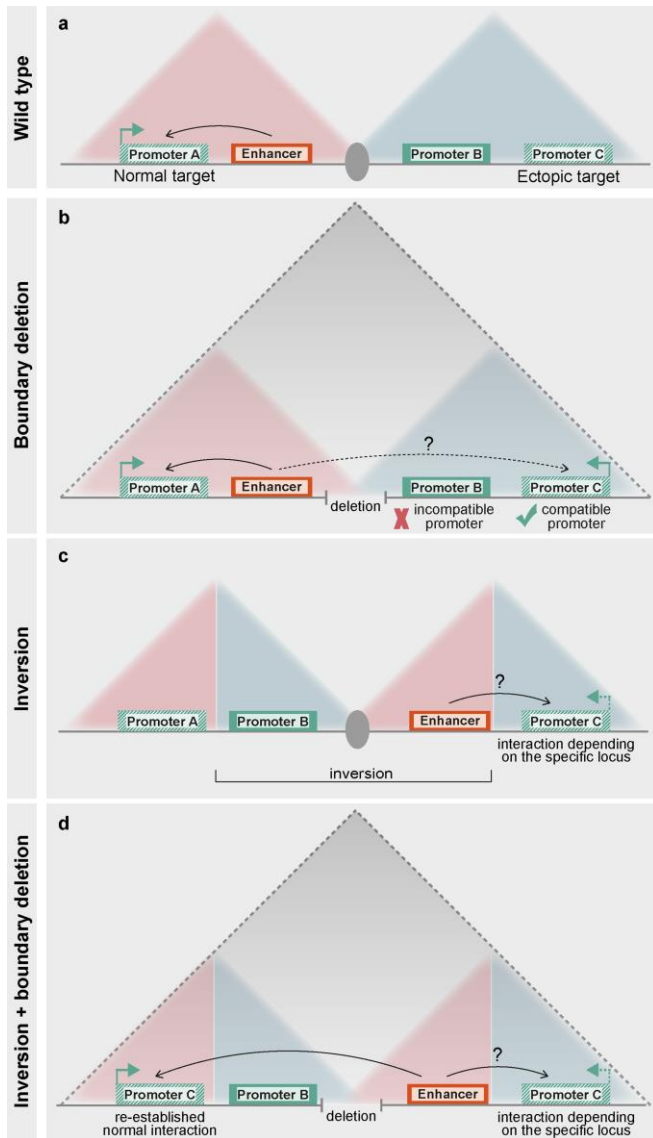


**Figure 7: Regulatory elements in enhancer-promoter communication.** Intra-domain enhancer-promoter interactions can take place in the absence of TADs. In contrast, TAD formation is required for inter-domain enhancer-promoter interaction (Yokoshi et al., 2020).

*(A) Do TADs boundaries constrain enhancer-promoter interactions?*

The hypothesis that many enhancer-promoter interactions are located within the same TAD, leads to the concept that TAD boundaries insulate genes from the action of enhancers located in the nearby domain (Figure 8a). It has recently been shown that the deletion of a TAD boundary typically leads to the 'fusion' of two adjacent TADs (Figure 8a-b). The functional impact of this deletion on gene expression appears to vary depending on the locus. In most cases, TAD fusions did not affect the ability of an enhancer to regulate its normal target gene, suggesting that TAD domains are not required for most enhancers to function. However, in some cases, TAD fusions did allow the enhancer to find a new target within the new fused TAD, indicating that the boundary was constraining its activity (Figure 8b-d). For example, in

mammals, the deletion of convergent CTCF sites in the  $\alpha$ -globin domain causes an expansion of the interaction domain without affecting  $\alpha$ -globin expression (Hanssen et al., 2017). However, this causes an ectopic expression of genes across the original boundary, indicating that this particular domain restricts enhancer activation but is not required for the enhancer's normal function.



**Figure 8: TADs organization. A.** Most enhancers and promoters are located within the same domain. **B.** Deletion of the boundary (grey oval) causes the 'fusion' of two TADs. Some enhancers (red rectangle) can still regulate their normal target gene (promoter A green) (left). Some enhancers will also regulate an

inappropriate gene (promoter C) (right). In a fused TAD, not all the genes are activated as promoter B. **C.** An inversion can also fuse two TADs and translocate the enhancer further away from its normal target gene, with a boundary in between. This can lead to a loss of interaction between an enhancer and its 'original' target gene (right). **D.** Deleting the boundary element can re-establish enhancer communication with its target gene (promoter A) while diminishing contact with promoter C (i.e. the 'original' target gene competes for the enhancer away from alternative promoters) (Cavalheiro et al., 2021).

An example where the expression of a gene is affected by the deletion of a boundary occurs in *Drosophila*, specifically in the *ftz* domain. The deletion of its boundary decreases the expression of a gene (*Scr*) in the neighboring domain. This suggests that the *ftz* domain is not involved in the regulation of the *ftz* gene, but rather helps to form a loop between the distal *scr* enhancer and its promoter (Calhoun and Levine, 2003; Yokoshi et al., 2020; Cavalheiro et al., 2021).

The breakdown of part of a TAD domain and its rejoining with another TAD domain after inverting its orientation is called 'inversion'. Inversions can disrupt an enhancers' target gene expression through the changes in the orientation of insulator binding motifs (Figure 8c) (Cavalheiro et al., 2021). For example, in mammals, large inversions at TAD boundaries, which place the *Epha4* enhancer cluster into a neighboring TAD, can cause the loss of *Epha4* expression and misexpression of different genes. However, this phenomenon also seems to be locus-dependant. Indeed, at the human *TFAP2A* locus, inversions that place an enhancer in a new TAD do not lead to the misexpression of neighboring genes (Laugsch et al., 2019). Similarly, in *Drosophila*, balancer chromosomes, which contain many inversions, deletions, and mutations, do not lead to associated changes in gene expression during embryonic development (Ghavi-Helm et al., 2019). This indicates that the recruitment of a new gene target by an enhancer following genomic rearrangements is highly locus-dependent.



Moreover, to understand how chromatin interactions can be translated into transcriptional outputs, a new method has been developed to position an enhancer in different regions relative to a fixed promoter within a TAD domain. Using the *Sox2* promoter fused to a Green Fluorescent Protein (GFP) and the *Sox2* control region (SCR) enhancer in mouse embryonic stem cells, this study showed that different ectopic insertions of an enhancer within a TAD domain have a distance-dependant effect on the transcription level of the target gene. Indeed, the transcription level of GFP inside the TAD domain depends on the genomic distance to the ectopic enhancers and decreases with increased genomic distance between the enhancer and the promoter. This means that the activity of the SCR enhancer is constrained by TAD boundaries (Zuin et al., 2021).

#### *(B) Do TADs play a role in facilitating enhancer-promoter communication?*

The influence of TAD domains on enhancer-promoter communication can be considered one of the key components for understanding the relationship between chromatin organization and gene expression regulation. Although still under intense investigation, TADs are involved in the formation of protein-mediated "loops" that bring pairs of genomic sites located far apart along the linear genome into spatial proximity (Field and Adelman, 2020). Currently, two potential mechanisms could explain the role of TADs in facilitating enhancers to regulate their target gene. The first one is increasing the frequency of enhancer-promoter interactions by reducing the impact of large genomic distance. The second one is increasing the local concentration of specific or general TFs through the formation of `hubs`. According to (Yokoshi et al., 2020), live imaging of *Drosophila* embryos shows that large enhancer-promoter distances significantly decrease the level of gene activity by affecting the timing and the size of transcriptional bursting. Interestingly, bursting profiles were recovered when the distal enhancer and target promoter were flanked by insulator pairs. Furthermore, when one of the insulator elements was deleted or inverted, the enhancer activity was not affected, suggesting that insulators at the boundaries of TAD domains can facilitate enhancers to induce transcriptional bursting independently of TAD formation.

However, it is still unclear whether similar effects would be observed for long-range enhancer-promoter interactions.

In mammals, the *Shh* locus constitutes an ideal system to study long-range enhancer-promoter interactions. Indeed, *Shh* expression is regulated by different tissue-specific enhancers located across a 900 kb region. It has been demonstrated that increasing or decreasing the genomic distance using genetic deletion or duplication, within the TAD domain, has no impact on *Shh* expression. It was only upon the inversion of the TAD boundary that led to a change in genome folding. Indeed, at this locus, regulatory contacts were decreased in a distance-dependent manner, causing *Shh* expression loss and limb malformation (Symmons et al., 2016). This data indicates that the *Shh* TAD can mitigate enhancer-promoter distance to achieve full robust expression, while limiting a possible boundary inversion, by decreasing the distance between enhancer-promoter pairs. Therefore, despite TADs may help to solve large linear genomic distances, their role is not absolute. The location of enhancers and promoters within a TAD are important determinants of interaction frequency and transcriptional output.

As noted above, TAD domains are largely invariant across cell types, rendering these domains insufficient to explain the highly dynamic nature of enhancer activity in different cell states. Moreover, within a single TAD domain, enhancers and promoters can vary widely in activity level, arguing domain-wide regulation of gene expression. For example at the *Shh* locus, inserted regulatory sensors across the TAD region revealed that the ZRS enhancer has non-uniform activity within the domain, indicating that enhancer-promoter specificity depends on more than proximity within a shared TAD.

In summary, placing an enhancer into a new TAD through TAD rearrangements is often not sufficient for it to activate neighboring genes. This suggests that the functional outcome of changing genome topology will be in part triggered by the compatibility of enhancers for their cognate promoters (Cavalheiro et al., 2021). Therefore, many of the main questions regarding the molecular mechanisms of enhancer-

mediated gene activation, the dynamics, and the specificity of enhancer-promoter communication, have yet to be fully explained. Besides, another key component required to achieve a global understanding of regulatory communication is the identification of the determinants of enhancer-promoter interaction specificity.

## 2.5 Molecular mechanisms underlying enhancer-promoter interaction specificity

Both classic and genome-wide studies have highlighted the complex network of enhancers and promoters. The mechanisms by which enhancers communicate with promoters and how promoter target specificity is defined have however remained elusive. Several potential molecular mechanisms have been identified to explain the specificity of enhancer-promoter interactions:

- A) biochemical compatibility
- B) spatial architecture of chromosomes within the nucleus
- C) insulator elements
- D) chromatin environment
- E) Tethering elements

### *(A) Biochemical compatibility*

Genetic studies demonstrated that some enhancers could interact with any promoter in their vicinity, whereas others prefer particular promoter types such as TATA- (core-promoter element), DPE- (downstream promoter element), or Inr- (Initiation element) containing promoters. This mechanism is called **biochemical compatibility**, which is the intrinsic ability of an enhancer-promoter pair to engage in a specific interaction through the recruitment of the right combination of proteins, which in turn is encoded in the DNA sequences of the enhancer and promoter (van Arensbergen et al., 2014). A good example is given by the regulation of *gsb* and *gsbn*, which are two divergently transcribed genes in *Drosophila* that have different expression patterns. Their enhancers GsbE and GsbnE are both located

around 10 kb away from their target promoters. Each enhancer could thus potentially cross-activate the other promoter, however, they do not. GsbE can only regulate the *gsb* promoter, whereas GsbnE can only regulate the *gsbn* promoter. Therefore, these enhancers exhibit distinct promoter compatibilities (Li and Noll, 1994). Another example of biochemical compatibility is the IAB5 enhancer of the Bithorax complex. In transgenic embryo assay, IAB5 preferentially activated a TATA-containing promoter but can also activate a DPE-containing promoter only if no TATA-containing promoter is available. Interestingly, however, at the endogenous locus, neither of the genes that are proximal to the IAB5 enhancer (*AbdA* and *AbdB*) have a TATA box. The selective activation of *AbdB* by IAB5 was found to be dependent on a 255bp element located in the proximal promoter. It is, therefore, an interesting possibility that different types of core promoter recognition complexes play a role in determining enhancer-promoter biochemical compatibility (Akbari et al., 2008). However, another study has shown that 14 out of 18 enhancers tested do not show specificity for DPE, Inr, or TATA motifs (Butler, 2001). Thus, core-promoter specificity could be a special property for a subset of transcriptional enhancers (Figure 9a) (Ohtsuki et al., 1998a; van Arensbergen et al., 2014).

Moreover, in *Drosophila*, the genes *decapentaplegic* (*dpp*), *SLY1 homologus* (*Slh*), and *out at first* (*oaf*) have been studied to understand why enhancers can regulate only specific promoters. It has been shown that *Slh* and *oaf*, which are located close to the *dpp* enhancers, are not affected by these *dpp* elements. Furthermore, when a transposon is located within the *oaf* gene, the *dpp* enhancers activate the more distant promoters while still ignoring the closer *Slh* and *oaf* genes. To test whether this promoter specificity accounts for autonomous gene regulation, *in vivo* gene targeting has been used to replace the *oaf* promoter with a *dpp*-compatible one. Strikingly, this ectopic gene has been activated by the *dpp* enhancers. These results suggest that the *dpp* enhancers preferentially interact with specific promoters. These enhancers do not activate the promoters of nearby genes but can activate a compatible promoter

inserted beyond them. A possible explanation, however, is that the activity of the *dpp* enhancers results from an altered chromatin environment. In this agreement, the insertion of the transposon in the genome may disrupt a chromatin structure that normally divides the *dpp* and *Slh/oaf* genes into separated domains. This chromatin perturbation would then allow the *dpp* enhancers to inappropriately activate any promoter in their proximity (Merli et al., 1996).

### *(B) Spatial architecture of chromosomes within the nucleus*

If the entire genome was randomly and homogeneously dispersed in the nuclear space, it would be nearly impossible for an enhancer to find a cognate promoter. Indeed the **spatial architecture of chromosomes within the nucleus** plays an important role in enhancer-promoter interactions. Enhancer and promoter can create physical contacts only if the overall folding of the chromatin fiber enables it.

TADs represent an architectural feature that can help distal enhancers to interact with the right target promoters (Figure 9b) (van Arensbergen et al., 2014; Cubeñas-Potts et al., 2017). Alteration of TAD boundaries can restructure the neighbouring TADs allowing an enhancer to create a contact and activate a gene located in an otherwise inaccessible TAD. This indicates that spatial chromatin conformation can facilitate the enhancer and promoter to find each other. However, not all promoters and enhancers within a TAD interact with each other, showing that there exists another layer of enhancer-promoter selectivity (Panigrahi and O'Malley, 2021).

### *(C) Insulator elements*

An alternative explanation for enhancer-promoter specificity is the presence of insulator elements. **Insulator elements** are DNA elements, bound by specific DNA-binding factors. Insulator elements can either act positively by facilitating enhancer-promoter interactions, allowing the regulation of a gene, or negatively by blocking essential enhancer-promoter contacts (Figure 9c) (van Arensbergen et al., 2014).

The first function of these elements is to prevent the activation of a promoter by an enhancer when placed between them. *Drosophila* has been a particularly good model system in which to analyse insulator function (Gurudatta and Corces, 2009). In *Drosophila* there are several types of insulators such as CTCF, BEAF-32, CP190, Zw5, Pita, ZIPIC, Ibf1/2, Su (Hw), Opbp. Each of these insulators consists of a DNA sequence that enrich specific DNA-binding proteins. Several experimental approaches and computational models showed that two or more insulator sites physically interact with each other, forming loops that alter the 3D conformation of the chromatin fiber in a manner that influences the ability of enhancers to interact with promoters (van Arensbergen et al., 2014). Depending on the specific context, such interactions can either block or facilitate enhancer-promoter contacts. For example, a recent study found that the *Drosophila* insulator Gypsy can insulate the *white* gene from the eye enhancer in part by directly interacting with it. It was observed that a specific component of the gypsy insulator interacted with the Zeste protein, which is critical for the eye enhancer-white promoter communication (Kyrchanova et al., 2013). There are many examples of enhancer-blocking activity of insulators, but most of these experiments have been performed in artificial ectopic settings, leaving the question of how relevant the effects identified are for endogenous function.

However, not all CTCFs exhibit this insulator activity (Lupiáñez et al., 2015). Indeed, the second function of insulators elements is that they can mediate enhancer-promoter interactions through their binding on promoter sequences. This mechanism is called CTCF binding affinity on promoter sequences. Recently it has been demonstrated that deletion of promoters containing CTCF-binding sites can cause the release of the partner enhancer, which loops and activates an alternative promoter (or alternative promoters) in the neighbouring regions. In particular, the enhancer exhibits a 'preference' in choosing one major target promoter to confer the strongest activation. Such promoter preference is defined by inherent features of the promoters and, sometimes, by the level of CTCF binding. These results recapitulate the 'enhancer scanning' model as the basis of enhancer-promoter engagement inside a contact domain (Blackwood and

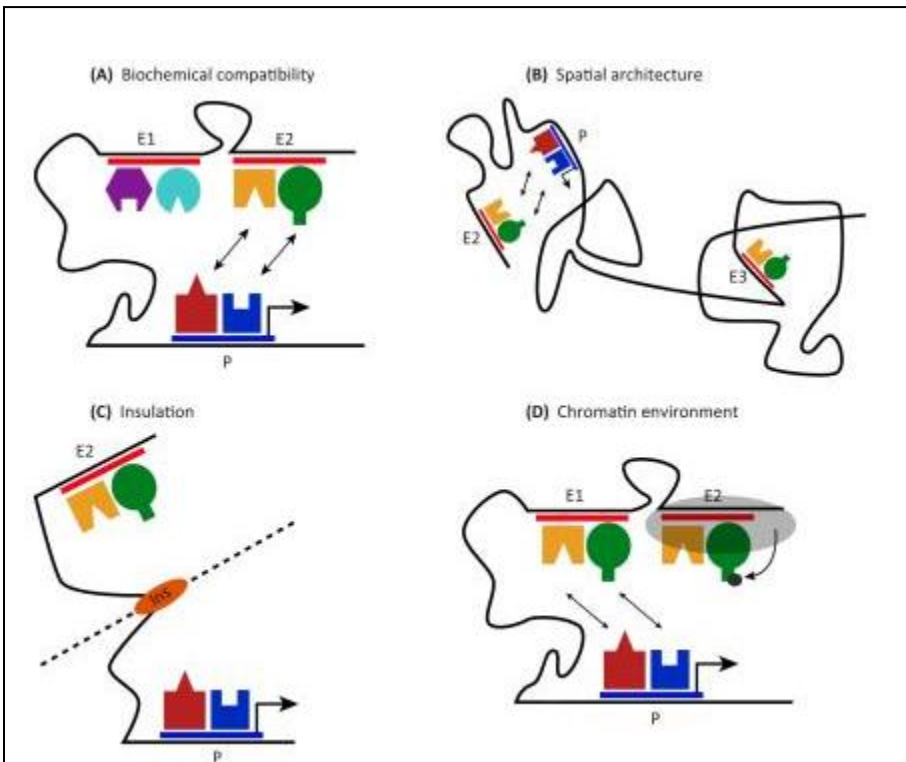
Kadonaga, 1998; Oh et al., 2021). Furthermore, because CTCF binding is observed at a substantial subset of promoters, it represents a mechanism for determining a subset of promoter preference events (Oh et al., 2021).

A recently developed technique with sub-kilobase resolution termed Micro-C has revealed short-range chromatin contacts that can occur in a CTCF dependent or CTCF independent manner and appear more cell-type-specific than traditional TADs (Hsieh et al., 2020; Krietenstein et al., 2020). Therefore, higher-resolution assays may provide new information on enhancer-promoter interactions. These studies highlight that there is still much to be learned and that CTCF mediated TADs are generally not sufficient to explain enhancer specificity nor are necessary for the maintenance of enhancer activity.

#### *(D) Chromatin environment*

Another mechanism that can contribute to the mutual selectivity of enhancer and promoter is the **local chromatin composition**. Approximately one-third of the genome is packaged into large chromatin domains that interact with the nuclear lamina. These lamina-associated domains (LADs) overlap partially with a subset of TADs. Most endogenous genes located in LADs show no detectable transcription and reporter assays showed that promoters ectopically inserted into LADs are less active than when integrated into inter LAD regions. Because this apparent chromatin effect on the ectopic promoters does not fully explain the complete lack of transcription, an interesting explanation is that the chromatin composition or spatial architecture of LADs further suppresses enhancer-promoter interactions (van Arensbergen et al., 2014). Although the effect of the chromatin context on enhancer-promoter interactions is still an active area of research, it is clear that local chromatin composition can affect enhancer and promoter activity. Indeed, tissue-specific enhancers are generally only active when bound by P300, and the H3K27ac chromatin mark distinguishes active enhancers from inactive enhancers containing H3K4me1 alone (Figure 9d) (van Arensbergen et al., 2014). Editing of histone modifications on enhancer sequences using

Transcription activator-like effector repeat domains (TAL), which are DNA binding domains that can target and disrupt DNA sequences, caused the specific downregulation of proximal genes, highlighting the importance of the chromatin composition of endogenous enhancers (Mendenhall et al., 2013).



**Figure 9: Mechanisms that can drive promoter-enhancer specificity.** **A.** Biochemical compatibility. Enhancer E2, but not E1, is compatible with the target promoter (P), indicating the selective interaction of E2 with P. **B.** Spatial architecture. Spatial architecture allows that only E2 interacts with the promoter, whereas E2 and E3 are both biochemically compatible with the promoter. **C.** Insulation. Although E2 is compatible with the promoter, their interaction is blocked by the insulator (Ins), possibly due to the alteration of the 3D structure. **D.** Chromatin environment. P specifically interacts with E1 because probably the compatibility of E2 with P has been altered by its chromatin environment (van Arensbergen et al., 2014).



### (E) Tethering elements

In *Drosophila*, promoter **`tethering` elements** can influence the specificity of enhancer-promoter interactions (Akbari et al., 2008; Calhoun and Levine, 2003). Such tethering elements have been studied in the bithorax complex of *Drosophila*. The IAB5 enhancer is located 55 kb downstream of the *abdominal B* promoter and 48kb upstream of the *abdominalA* promoter. Interestingly, the IAB5 enhancer specifically interacts only with the *abdominalB* promoter, even though the enhancer and promoter are separated by at least two insulators. This occurs due to the presence of a 255bp element, located upstream of the *abdominalB* transcription start site. This element, called tethering element, can help the interaction between the IAB5 enhancer and the *abdominalB* promoter in transgenic embryos (Akbari et al., 2008). This evidence shows that a novel class of *cis*-regulatory elements might regulate enhancer-promoter interactions specificity.

To achieve a deep understanding of the complex network of enhancer-promoter interactions, large-scale functional studies will be essential. In particular, systematic approaches combined with computational analyses will lead the way towards a better insight into the determinants of enhancer-promoter interactions. Scalable approaches in which functional enhancer-promoter interactions are perturbed in a targeted manner may help to establish these determinants. For example through the dissection of specific loci and precise live monitoring of chromatin organization and gene expression in single cells.

### **3. *Drosophila melanogaster* early embryogenesis and mesoderm specification**

#### **3.1 Early *Drosophila* embryogenesis**

In the past decade, studies have revealed that the genetic cascades driving development are highly conserved between the fruit fly *Drosophila melanogaster* and vertebrates. Therefore, *Drosophila* has become an essential model organism in developmental biology with many genetic and genomic tools available. In particular, its relatively small thoroughly annotated genome (120 Mb), and very well-characterized developmental gene regulatory network, make *Drosophila melanogaster* a model of choice for the study of chromatin organization and enhancer-promoter interactions during embryogenesis (Moretti et al., 2019; Rivera-Pomar and Jäckle, 1996). Furthermore, advances in the genetic technologies available to scientists studying *Drosophila* have allowed a wealth of knowledge on the developmental regulation of gene expression, tissue patterning, enhancer biology, and enhancer-promoter interactions (Jennings, 2011).

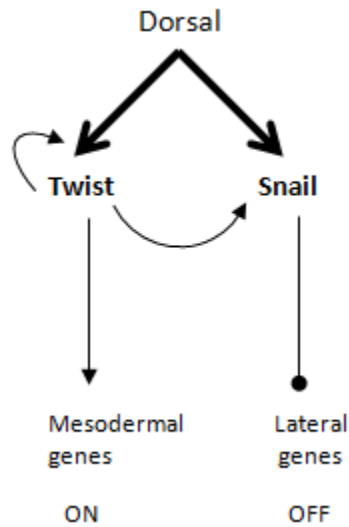
*Drosophila* embryos develop very rapidly, within about 18 hours from egg-laying to completion of embryogenesis. Following fertilization, the zygote nucleus rapidly divides in 13 synchronous mitotic cycles, resulting in a multi-nucleated cell called the syncytial blastoderm. The nuclei then migrate to the periphery of the embryo and cellularise, forming the cellular blastoderm. By this time, the zygotic genome, which was previously quiescent, starts being transcribed in a process called maternal to zygotic transition. After the formation of the cellular blastoderm, about 3h after egg-lay (AEL), gastrulation begins, which gives rise to the three primary germ layers: mesoderm, endoderm, and ectoderm. Interestingly, by the onset of gastrulation, cell fates are already specified at a single-cell level.

The origin of the embryo body axes is determined well before fertilization by the deposition of maternal mRNA molecules. When these molecules are translated into proteins called morphogens, a concentration gradient is established leading to the patterning of the early embryo along its anterior-posterior and

dorso-ventral axes (Thisse et al., 1991).

### **3.2 Mesoderm specification**

The maternal gradient of the *dorsal* transcription factor, which is homologous to NF-kappaB, defines the different regions of the embryo along the dorsal-ventral axis. Dorsal controls the dorsal/ventral pattern by activating the Twist and/or Snail transcription factors, which are required for the differentiation of the mesoderm. At the molecular level, the maternal function of *dorsal* is required for *twist* expression. Indeed, *twist* RNA is not detected in embryos derived from eggs laid by mutant *dorsal* females. In particular, Dorsal acts as a sequence transactivator of the *twist* promoter, and *in vitro* analysis has shown that Dorsal binds *twist* regulatory sequences. *twist* and *snail* mutants gastrulate abnormally, form no mesodermal germ layer or mesoderm derivatives, and die before the end of embryogenesis. It has been shown that *twist* responds to maternal positional information defining the boundaries of the mesoderm. Within this territory, it is responsible for activating and maintaining the expression of early mesoderm-specific genes, including itself and *snail*. However, these two genes have distinct roles in the formation of the mesoderm. Indeed, *twist* is necessary for the activation of downstream mesodermal genes and *snail* prevents the expression of genes in the mesoderm, which are normally active only in more lateral or dorsal regions (Figure 10). The absence of *twist* and *snail* results in the complete loss of all mesodermal characteristics (Leptin, 1991).



**Figure 10: Diagram summarizing early gene interactions during mesoderm specification.** Large black arrows represent maternal regulation of the early *twist* and *snail* transcription. Thin arrows denote activation. The line ending with a black dot denotes repression.

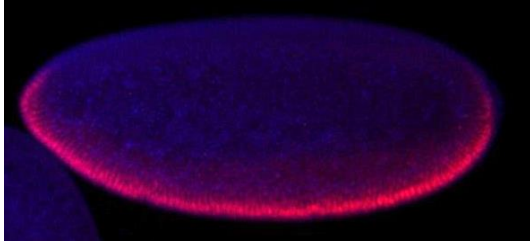
The formation of muscles during embryonic development is a complex process that requires the coordinated actions of many genes. Somatic, visceral, and heart muscles are all derived from mesoderm progenitor cells. The *twist* gene, which encodes a bHLH transcription factor, is essential for multiple steps of mesoderm development: invagination of mesoderm precursors during gastrulation, segmentation, and specification of muscle types (Furlong et al., 2001). In addition to its role in mesoderm formation, Twist has a second function in later stages after gastrulation, where it appears to be required for the myogenesis of somatic muscles (Yin et al., 1997). There are two ways in which *twist* could function in somatic myogenesis. During early embryogenesis, *twist* is expressed in a modulated fashion in which progenitors of somatic muscles contain a relatively large amount of *twist* expression, whereas a small amount of expression occurs in the visceral mesoderm and the heart. Later, *twist* expression is lost from cells as they differentiate to form larval muscles but it remains in cells that will proliferate to form the adult myoblasts. These two features of *twist* expression suggest that there are two alternative moments when *twist* is

important in the mesoderm. **(1)** The *twist* gene has an early function in the subdivision of mesoderm into different derivatives. **(2)** The *twist* gene has a late function in switching cells between proliferating and differentiating states in the somatic myogenic lineage (Baylies and Bate, 1996).

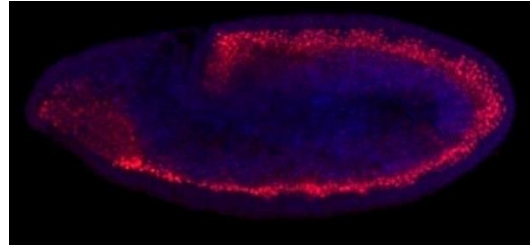
### **3.3 *twist* regulation during early embryogenesis**

To identify *in vivo* sequences necessary to rescue the *twist* mutant phenotype and the regulatory sequences required for the expression of *twist*, different phenotypic rescue analyses and *twist*  $\beta$ -galactosidase expression fly lines have been used. Through P element transformation, different regulatory regions upstream and downstream of the *twist* gene have been studied to analyze the phenotypic rescue in mutant animals until adulthood. As a result, a 9 kb fragment (-3 kb to 6.2 kb) was found to rescue all *twist* alleles and includes all the regulatory elements required for the correct spatial and quantitative expression of *twist*. This indicates that the 9kb fragment spanning the *twist* gene is indispensable for the phenotype of the embryo (Thisse et al., 1991).

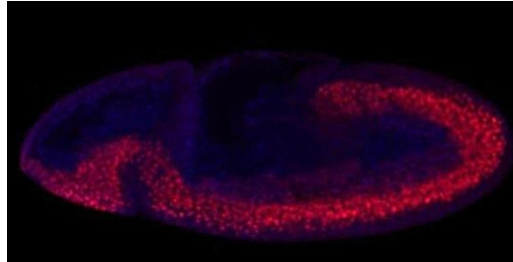
During early *Drosophila* embryogenesis *twist* is expressed mainly in the ventral side of the embryo at different stages (4-6, 7-8, 8-9 AEL) as shown in Figure 11. During early embryogenesis *twist* is regulated by three enhancers: an upstream distal enhancer DE (-1.2kb to 800bp), an upstream proximal enhancer PE (-440bp to -180bp), and a downstream enhancer VT21150 (about 1kb-long) (Jiang et al., 1991; Kvon et al., 2014; Pan et al., 1991; Thisse et al., 1991). For simplicity, we will refer to them as enhancer E1, E2, and E3 respectively. Those enhancers are located in open chromatin regions and are bound by mesodermal TFs such as Twist, Tin, and Mef2, as well as insulator proteins (Figure 12) (Nègre et al., 2010; Thomas et al., 2011; Zinzen et al., 2009). As expected, the spatial activity of *twist* enhancers occurs in the most ventral region of *Drosophila* embryos.



stage 4-6



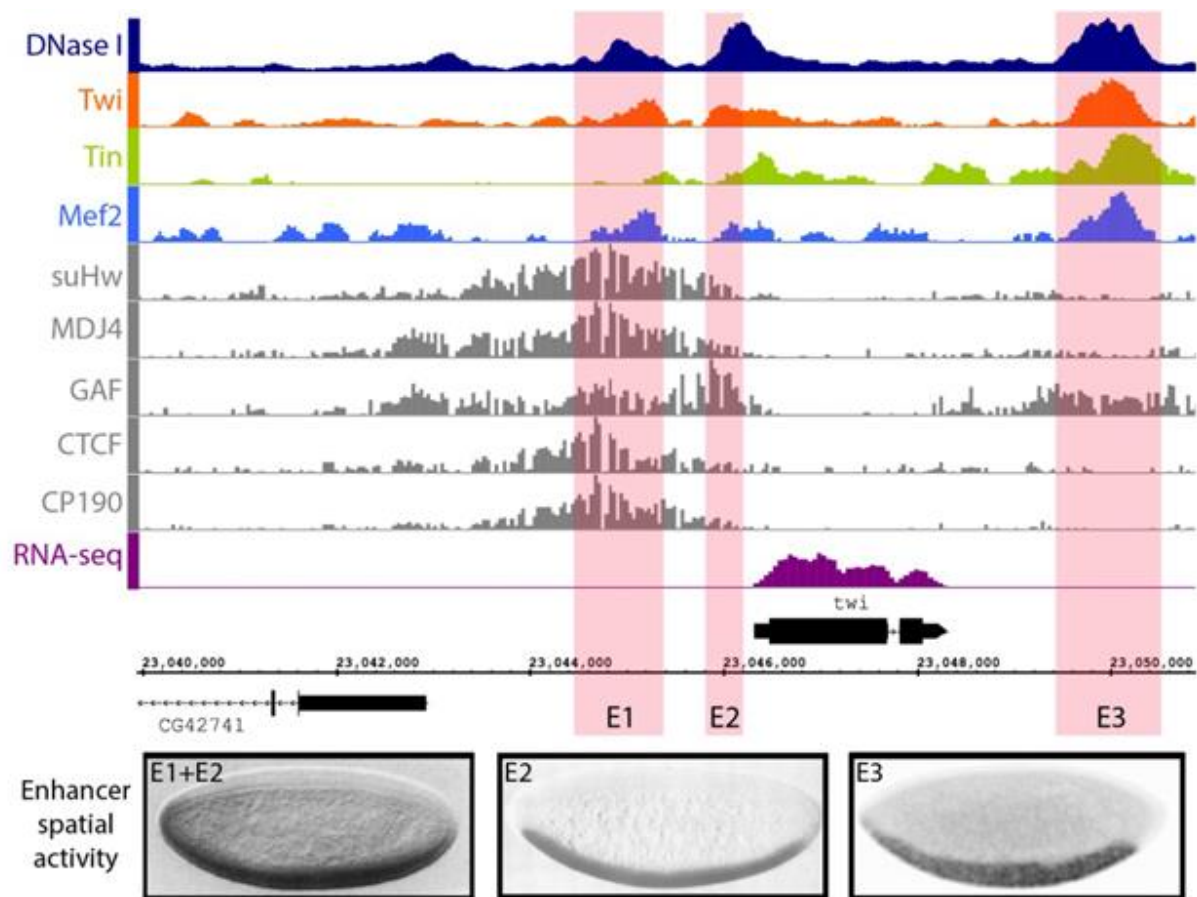
stage 7-8



stage 8-9

**Figure 11: Twist expression at three different time points during Drosophila embryogenesis.**

Immunostaining from 2-4h AEL (stage 4-6 and 7-8) until 4-6h AEL (stage 8-9).



**Figure 12: The Twist locus and its three enhancers. Top:** Dnase-seq at 2-4 AEL, followed by ChIP-chip signal at 2-4 AEL for Twist (orange), Tinman (green) and Mef2 (light blue). ChIP-chip signal at 0-12h AEL for insulator proteins such as suHw, MDJ4, GAF, CTCF, CP190. The location of the three twist enhancers is highlighted. RNA-seq signal at 2-4h AEL (purple) (Graveley et al., 2011; Nègre et al., 2010; Thomas et al., 2011; Zinzen et al., 2009). **Bottom:** Activity driven by the *twist* enhancers at 2-4 AEL (immunostaining) (Jiang et al., 1991; Pan et al., 1991; Thisse et al., 1991).

#### 4. Concluding remarks

Although many studies have contributed to understanding the relationship between chromatin organization and enhancer-promoter interactions, the molecular mechanisms driving the interaction

specificity between enhancers and promoters remain unknown. As enhancers were shown to be the major regulators of developmental gene expression, understanding their properties and their functions in relation to chromatin domains is essential. During my Ph.D. I used the regulation of *twist* with its E3 enhancer, to provide new insights into the role of the genomic context in mediating enhancer-promoter interaction specificity and in mesoderm development during early *Drosophila melanogaster* embryogenesis.



# Aims

In *Drosophila*, chromatin interactions of distal and proximal regulatory elements such as enhancers and promoters seem to be mostly constrained throughout an entire TAD domain rather than neighboring TADs. This indicates that TADs might promote interactions between enhancers and promoters. While a given enhancer could, in theory, interact with all the promoters located within its TAD domain, it actually interacts with just a subset of them, conferring a layer of specificity between enhancers and promoters. Therefore, several molecular mechanisms must be involved to establish the affinity of an enhancer for specific promoters. The mechanisms governing this specificity are still largely unknown and are the focus of my Ph.D. project.

My Ph.D. project aims at deciphering how the genomic context influences enhancers-promoters interactions. This will help us to understand how enhancers choose a specific promoter within the context of TAD domains. Using the *twist* gene of *Drosophila melanogaster* as a model, I systematically perturbed enhancer-promoter interactions in a controlled manner to assess the role of the genomic context on their interaction. In particular, I used the most distal enhancer of *twist* (referred to as E3) to answer the following question: does the genomic context influence enhancer-promoter interactions?

My Ph.D project is divided in three aims:

1. To test if the enhancer E3 functions in an orientation-independent manner at the endogenous locus, I inverted the orientation of enhancer E3 sequence and I assessed the impact on chromatin interactions and on the expression of *twist*
2. To test if the chromatin environment can influence chromatin interactions between enhancer E3 and the *twist* promoter, I ectopically inserted enhancer E3 in different genomic locations that differ from its endogenous locus, both in terms of distance and chromatin environment.

3. I then tested how these ectopically inserted E3 enhancers could influence the transcription of nearby genes.

Taken together these experiments provide a better understanding of enhancer biology, chromatin organization, transcription regulation and will contribute to the understanding of specificity between enhancers and promoters during early *Drosophila* embryogenesis.

# Chapter 2: Results



## 1. Introduction

Embryonic development is driven by precise patterns of gene expression, directing a cell to establish a particular fate. These expression patterns are mostly regulated by the action of *cis*-regulatory elements, such as promoters and enhancers. Enhancers are short regulatory elements located in the non-coding part of the genome containing DNA motifs, which act as binding sites for transcription factors and that can drive the expression of a reporter gene in an orientation- and context-independent manner when ectopically inserted into the genome. Enhancers can sometimes be located at great linear distances from the gene that they regulate (Lettice et al., 2003), on average 120 kb away in mammals (Jin et al., 2013; Sanyal et al., 2012), often skipping intervening genes (van Arensbergen et al., 2014) or even been located on different chromosomes (Lomvardas et al., 2006). When placed at large genomic distances, enhancers can contact the promoter of their target gene(s) via the formation of a chromatin loop (Krivega and Dean, 2012). Even in the compact *Drosophila* genome, the distance between enhancers and their target genes is very comparable to mammals, with a median distance of 100 kb, and some interactions even spanning distances of over 500 kb (Ghavi-Helm et al., 2014). Importantly, the enhancers' target genes are not necessarily the closest genes. Only 27% of the enhancer-promoter interactions identified in human cell lines are between neighboring elements, and this number increases only up to 47% when considering solely the nearest active promoter (Sanyal et al., 2012). In this context, it is essential for enhancers to target and regulate the expression of the correct gene while avoiding the inappropriate expression of other neighboring genes.

In recent years, topological constraints of the genome have been suggested to play an important role in mediating enhancer-promoter interactions. Indeed, most enhancer-promoter interactions happen within large domains of increased three-dimensional proximity, named topologically associating domains (TADs), which broadly correspond to regulatory domains and are enriched in co-regulated genes (Le Dily et al., 2014; Nora et al., 2012; Shen et al., 2012; Symmons et al., 2014a; Zhan et al., 2017). In addition,

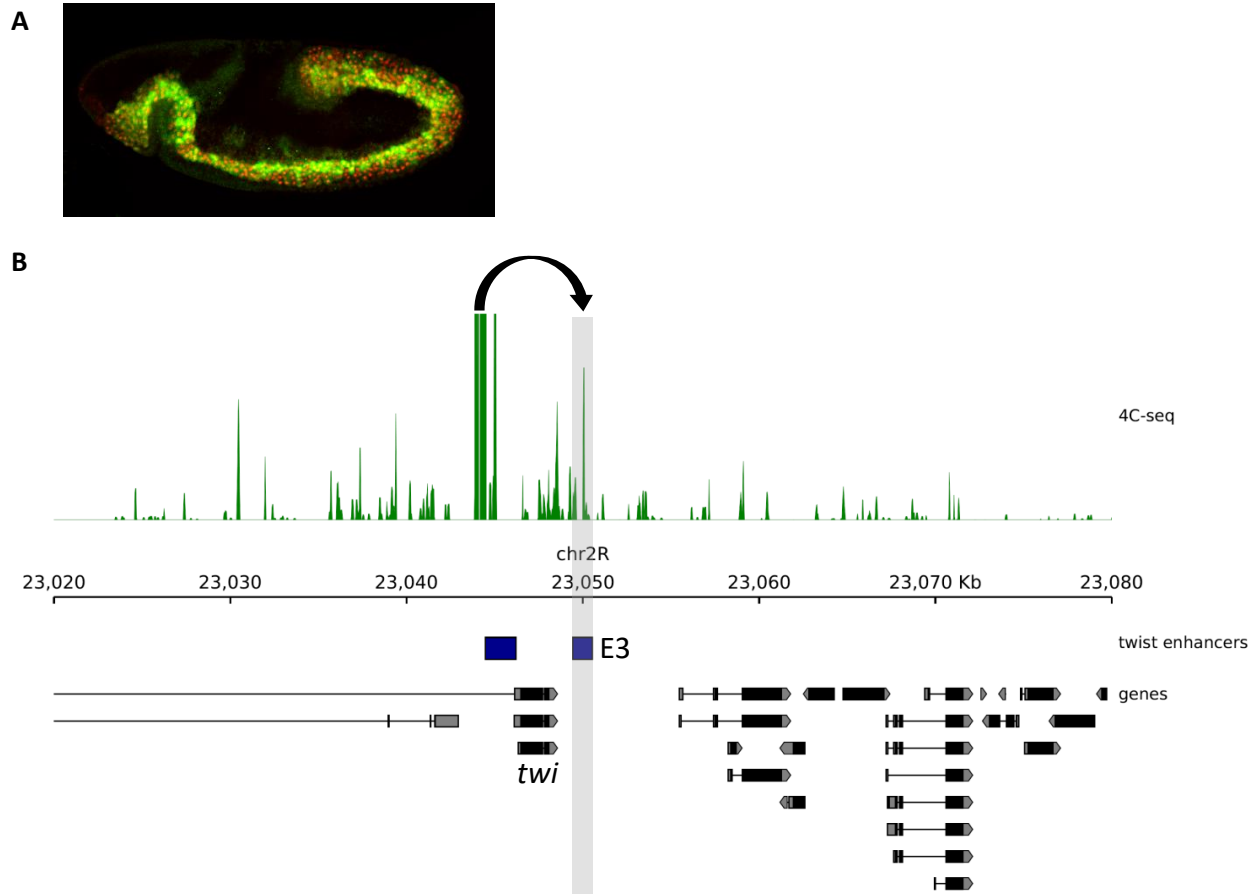
chromosomal rearrangements such as inversion, deletions, or duplications that disrupt TAD boundaries often impair enhancer-promoter communication and cause the misexpression of genes (Flöttmann et al., 2015; Franke et al., 2016; Giorgio et al., 2015; Lupiáñez et al., 2015b; Redin et al., 2017; Spielmann et al., 2012). TADs might thus restrain enhancer-promoter interactions by either favoring interactions within the same TAD or preventing remote interactions across TAD boundaries. However, recent evidence contradicts this “TAD-centric” view. First, the depletion of CTCF and cohesin, which are key players in the formation of TAD boundaries, totally abolished TAD structures, but only mildly affect gene expression (Nora et al., 2017a; Rao et al., 2017; Schwarzer et al., 2017). Second, chromosomal rearrangements that affect TAD boundaries don’t necessarily affect gene expression (Despang et al., 2019; Dong et al., 2018; Ghavi-Helm et al., 2019; Laugsch et al., 2019). For example, in the highly scrambled *Drosophila* balancer chromosomes, changes in chromatin organization do not directly impact gene expression (Ghavi-Helm et al., 2019). While chromatin organization is undoubtedly a major player in targeting an enhancer to the correct promoter, enhancer-promoter interaction specificity cannot be explained by genome topology alone (Ghavi-Helm, 2019).

Apart from TADs, several other factors are involved in mediating enhancer-promoter specificity (van Arensbergen et al., 2014). These include the biochemical compatibility of a given enhancer for particular promoter motifs (Li and Noll, 1994; Ohtsuki et al., 1998b), the presence of insulator elements that can either act as enhancer-blocking elements (Kyrchanova et al., 2013) or promote enhancer-promoter interactions, the local chromatin composition of the enhancer-promoter pair, or the presence of tethering elements (Calhoun et al., 2002). Overall, enhancers and promoters must engage in specific interactions to ensure the correct spatio-temporal expression of genes. While it is likely that several factors act in combination to determine enhancer-promoter interaction specificity, the precise rules governing this mechanism remain elusive.

To address this question, we used genome engineering to generate a number *D. melanogaster* fly line where one of the enhancers of the *twist* gene, the master regulator of mesoderm fate, has been misplaced at different genomic locations. These fly lines were used to assess how the location of an enhancer can affect its ability to interact with its target gene and drive faithful gene expression.

## **2. Using the *twist* E3 enhancer as a model to study enhancer-promoter interactions**

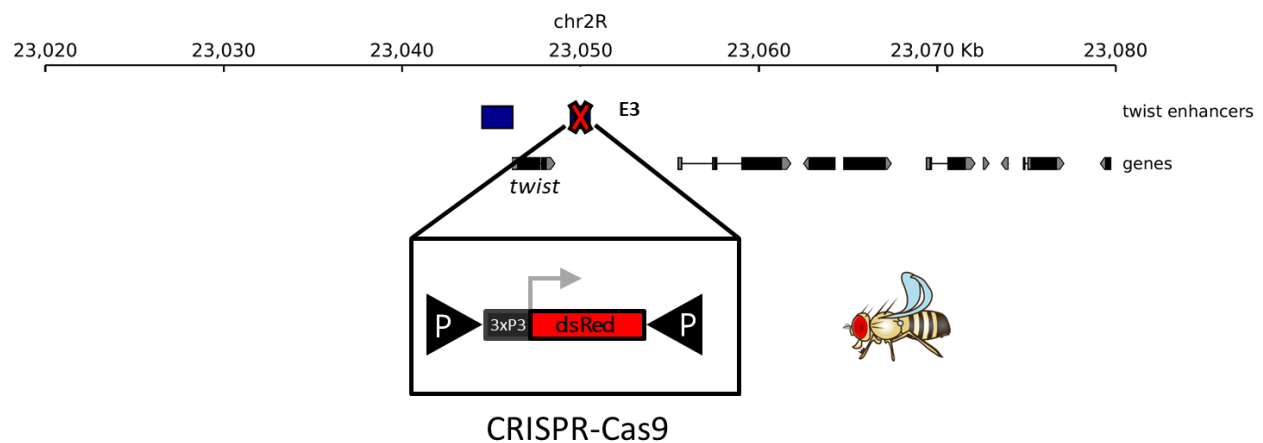
During early embryogenesis, the expression of the *twist* gene is regulated by three enhancers, which will hereafter be referred to as E1, E2, and E3 (Kvon et al., 2014; Pan et al., 1991; Thisse et al., 1991). The E3 enhancer, which is located about 3 kb downstream of the *twist* promoter, recapitulates the expression of the *twist* protein during the early stages of embryogenesis in enhancer-reporter assays (Figure 13A). We first verified that the endogenous E3 enhancer interacts with the *twist* promoter at the stages when the gene is expressed. For this purpose, we performed 4C-seq (chromosome conformation capture sequencing) experiments in *y*[\*], *w*[\*] (wild-type) *Drosophila* embryos, using a viewpoint located 2kb upstream of the *twist* promoter. As expected, we observed an increased interaction frequency between the viewpoint and a region overlapping the E3 enhancer, confirming that the E3 enhancer can interact with the promoter of *twist* (Figure 13B).



**Figure 13: Characteristics of enhancer E3.** **A.** Immunostaining showing the overlap between the expression of the twist protein (red) and the expression of a GFP reporter driven by the E3 enhancer (green) at stage 8. **B.** 4C interaction map in wild-type over-night *Drosophila* embryos showing the expected interaction between the *twist* promoter (arrowtail; viewpoint) and the endogenous enhancer E3 (arrowhead; highlighted in grey).



To further characterize the role of the E3 enhancer, we used CRISPR-Cas9 homology-directed repair (HDR) to create a new fly line (hereafter referred to as “delta E3”) where the endogenous E3 sequence has been deleted (Figure 14). In this fly line, the endogenous E3 enhancer has been replaced by a *3xP3::dsRed* fluorescent cassette flanked by two inverted  $\Phi$ C31 integrase *attP* sites. This deletion causes a homozygote lethal phenotype at embryonic stages, indicating that the E3 enhancer is essential for proper embryogenesis and that its activity is not redundant with that of the upstream E1 and E2 enhancers. We are currently performing a more detailed phenotypic characterization of this mutant to assess potential muscle defects.

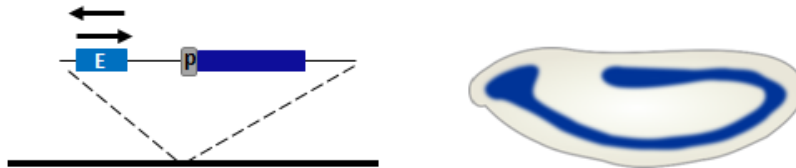


**Figure 14: The deletion of the endogenous E3 enhancer is homozygote lethal.** The endogenous E3 enhancer sequence was replaced by a *3xP3::dsRed* fluorescent cassette (driving red fluorescence in the eyes) flanked by two inverted  $\Phi$ C31 integrase *attP* sites. The resulting fly line is homozygous lethal and kept as heterozygous over a CyO balancer chromosome.

### 3. The E3 enhancer functions in an orientation-independent manner at its endogenous locus

The assumption that enhancers can activate the transcription of their target gene independently of their orientation and location is largely based on enhancer-reporter assays, where the enhancer, cloned directly

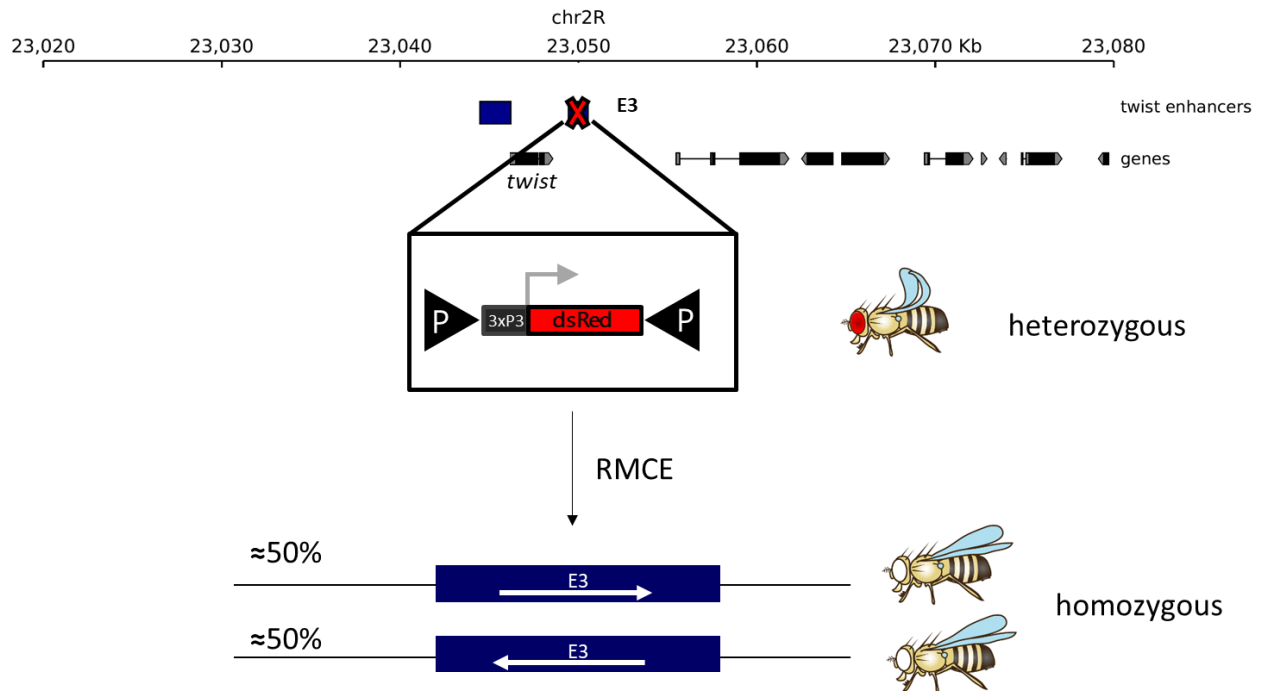
upstream of the promoter, does not require looping to drive gene expression (Figure 15). Whether the orientation of an enhancer at its endogenous locus could affect the interaction with its target promoter and/or the expression level of its target gene remains however elusive.



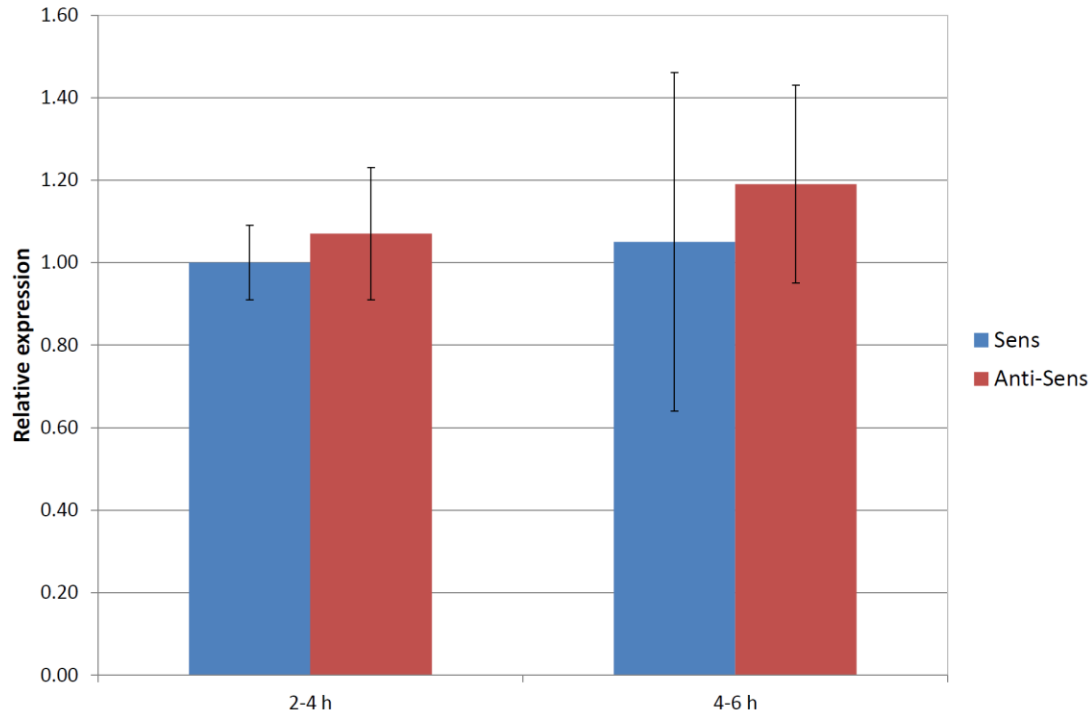
**Figure 15: Schematic example of an enhancer-reporter assay.** A construct containing an enhancer (E, light blue box) cloned upstream of a minimal promoter (P, grey box) and a reporter gene (dark blue box) is randomly integrated into the genome. The expression of the reporter gene (blue pattern in the embryo) is activated independently of the enhancers' orientation and recapitulates the activity of the enhancer.

We took advantage of the delta E3 fly line to test whether the orientation of the endogenous E3 enhancer could affect the expression of *twist*. We created a vector (*pBS-KS-attB-E3*) where the E3 enhancer is flanked by two inverted  $\Phi$ C31 integrase *attB* sites and used recombinase-mediated cassette exchange (RMCE) to replace the *3xP3::dsRed* fluorescent cassette with the E3 enhancer. With this strategy, the E3 enhancer can be re-integrated randomly in either orientation. Positive flies were screened for the loss of red fluorescence in the eyes, then genotyped by PCR to assess the orientation of the insertion. We obtained two fly lines where the endogenous E3 enhancer was re-inserted either in its original orientation (hereafter called "Sens") or in the opposite orientation (hereafter called "Anti-Sens"). In both cases, the flies are homozygous, demonstrating that the orientation of the enhancer at its endogenous locus does not cause any lethality (Figure 16). To assess if the enhancers' orientation could influence the expression of *twist*, we performed Reverse Transcription quantitative PCR (RT-qPCR) in *Drosophila* embryos collected

at two different time points (2 to 4 hours and 4 to 6 hours after egg-lay). This experiment revealed that the orientation of the endogenous E3 enhancer doesn't alter the expression of *twist*, at least during early embryogenesis (Figure 17). We then asked whether the chromatin loop present between the *twist* promoter and the endogenous E3 enhancer (Figure 13B) was affected. We collected *Drosophila* embryos from the Sens and Anti-Sense lines at 2 to 5 hours after egg-lay and performed 4C experiments using the same as previously (located upstream of the *twist* promoter). Unfortunately, this 4C-seq data was of slightly lower quality than the wild-type data presented in Figure 18. The E3 enhancer been less than 5 kb away from the *twist* promoter, 4C-seq experiments are particularly challenging at this locus. We observe, however, that the interaction seems maintained regardless of the enhancers' orientation (Figure 18), but the experiment would need to be repeated to be fully conclusive. Given the absence of effect on the expression of *twist* by RT-qPCR, we don't expect in effect on chromatin organization and thus decided to not repeat this experiment. Therefore, we conclude that the E3 enhancer functions in an orientation-independent manner even at its original endogenous locus.

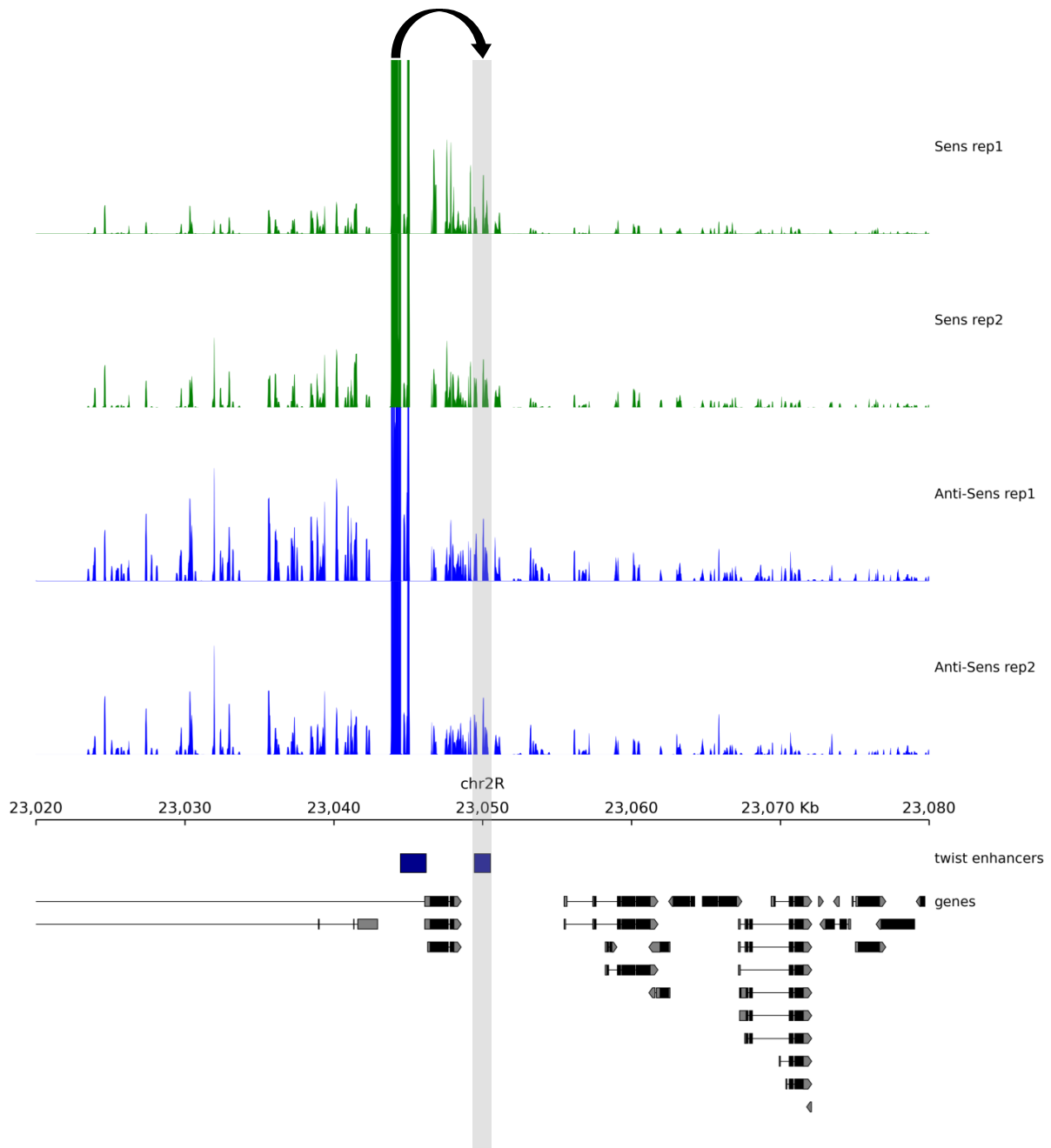


**Figure 16: Schematic overview of recombinase-mediated cassette exchange.** The *3xP3::dsRed* fluorescent cassette (driving red fluorescence in the eyes) was replaced by the E3 enhancer using RMCE. The enhancer is randomly re-inserted in either orientation. The resulting fly lines are homozygous viable.



**Figure 17: The expression of *twist* is not affected by the orientation of the endogenous E3 enhancer.**

Relative expression of the *twist* gene normalized over the housekeeping gene *RpL32* in *Drosophila* embryos from the Sense (blue) and Anti-Sens (red) lines collected at two different time points (2-4h and 4-6h after egg-lay). The error bar represents the standard deviation between two biological replicates.



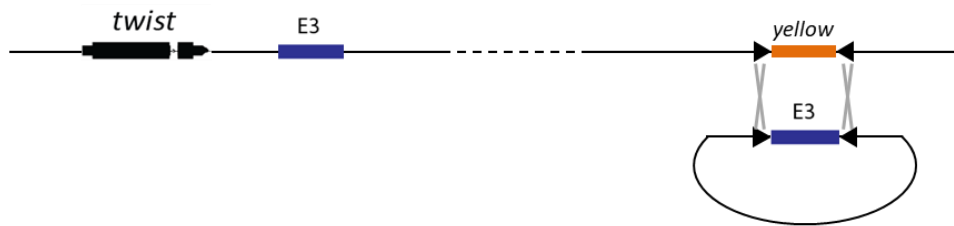
**Figure 18: The E3 enhancer probably interacts with the *twist* promoter irrespective of its orientation.**

4C interaction map in the Sens (green) and Anti-Sens (blue) lines. *Drosophila* embryos were collected at 2 to 5 hours after egg-lay in two biological replicates. The expected interaction between the *twist* promoter (arrowtail; viewpoint) and the endogenous enhancer E3 (arrowhead) is highlighted in grey. The location of known *twist* enhancers is indicated by blue rectangles.

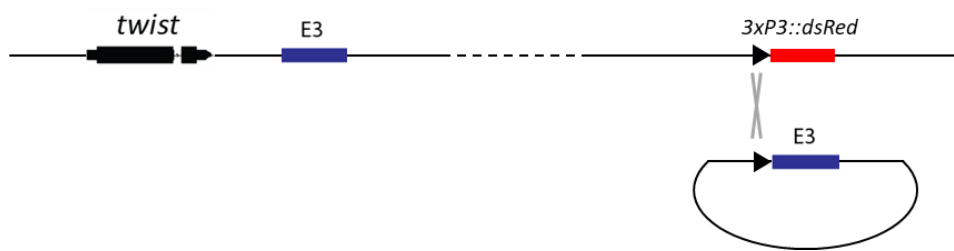
#### 4. Generating transgenic fly lines where the E3 enhancer has been inserted in various locations

We next wanted to test whether the location of an enhancer could influence the ability of this enhancer to regulate the expression of its target gene. For this purpose, we decided to generate several fly lines where the E3 enhancer of *twist* was inserted in different locations, at various distances from the *twist* promoter, and in regions characterized by various chromatin landscapes. To create these lines, we used two strategies (Figure 19): First, we took advantage of the popular MiMIC (*Minos* Mediated Integration Cassette) system (Venken et al., 2011), which consists of a *Minos* transposon carrying a *yellow*<sup>+</sup> dominant body-color marker and a gene-trap cassette flanked by two inverted  $\Phi$ C31 integrase *attP* sites. This cassette can be efficiently replaced by another cassette contained the DNA sequence of interest flanked by two inverted  $\Phi$ C31 integrase *attB* sites using RMCE. This insertion event can be conveniently identified by the loss of body pigmentation in adult flies (corresponding to the replacement of the *yellow*<sup>+</sup> marker by the sequence of interest). We used this strategy to generate five different fly lines where the E3 enhancer was inserted at different locations (Figure 19A) (see insertion coordinates and MiMIC fly lines in the Material and Methods section, table 9). While such MiMIC fly lines are readily available to create insertions at thousands of sites, we had to use a second strategy to specifically insert the E3 enhancer in the same TAD as *twist*. In this second strategy, we first used CRISPR-Cas9 mediated HDR to create a fly line where a  $\Phi$ C31 integrase *attP* site is integrated at the desired location, at coordinate 23,053,901 on chromosome 2R. This site was then used to insert the E3 sequence from a donor vector containing a  $\Phi$ C31 integrase *attB* site (Figure 19B). Importantly, the ectopically inserted E3 sequence was amplified from genomic DNA extracted from a different fly line than the MiMIC and Cas-9 expressing lines. Therefore, the endogenous E3 sequence can be differentiated from the ectopically inserted E3 sequence by the presence of several single nucleotide variants. Overall, we generated six fly lines where the E3 enhancer was inserted either upstream or downstream of the *twist* promoter, at several locations.

**Strategy 1: using MiMIC fly lines A**



**Strategy 2: using CRISPR-Cas9 HDR B**



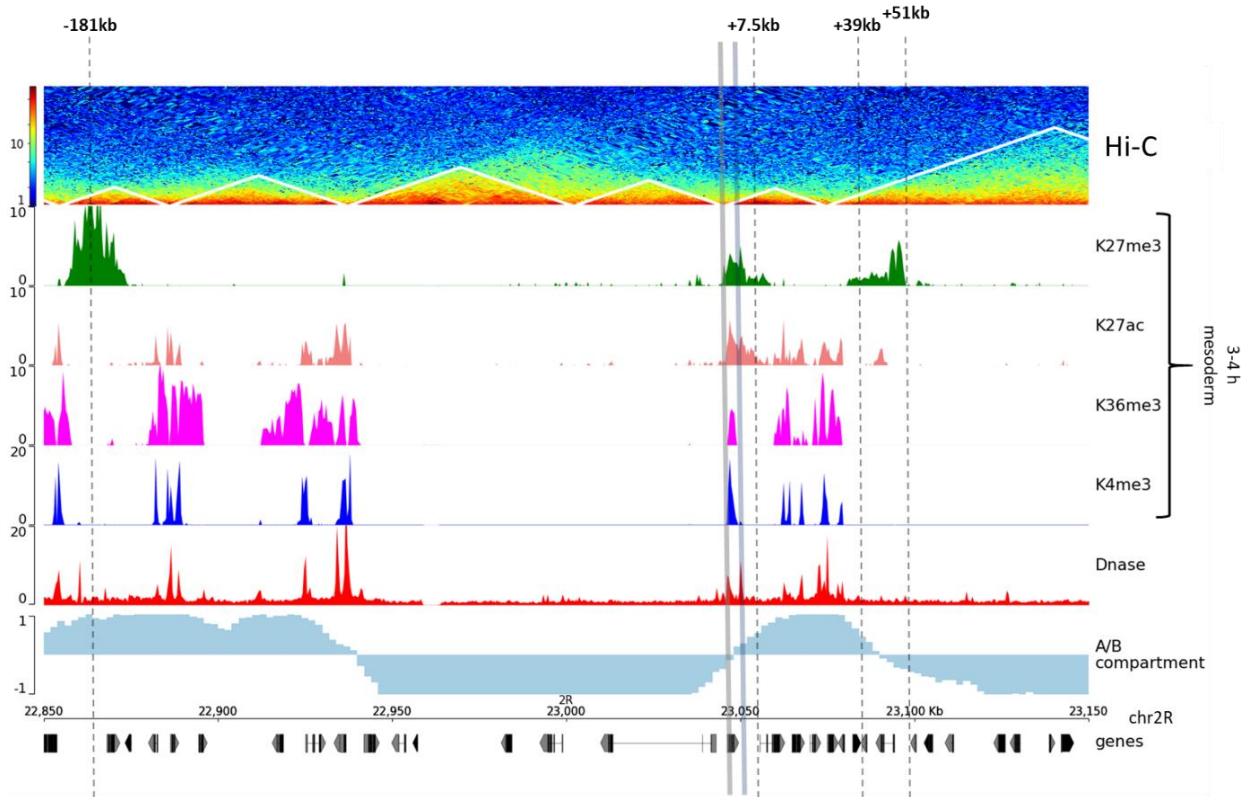
**Figure 19: Overview of the two different experimental strategies used in this study to insert the enhancer E3 at ectopic locations. A. One-step strategy.** The MiMIC construct consists of two  $\Phi$ C31 integrase *attP* docking sites (black triangles) flanking a *yellow*<sup>+</sup> marker (brown box). The *yellow*<sup>+</sup> marker is replaced by the E3 sequence (blue box) using RMCE. **B. Two-step strategy.** A  $\Phi$ C31 integrase *attP* site and a *3xP3::dsRed* (red box) marker site were inserted at the desired location using CRISPR-Cas9 HDR.  $\Phi$ C31 integrase-mediated recombination is used to insert the E3 enhancer from a vector containing a  $\Phi$ C31 integrase *attB* site.



The main characteristics of the different transgenic fly lines are summarized in Table 2. In addition, the insertion sites are located in regions that do not overlap DNase I hypersensitive sites during embryogenesis based on (Thomas et al., 2011). Furthermore, the insertion sites were selected based on their distance to the *twist* promoter (one site is located in the same TAD as *twist*, two sites are located in the next neighboring TAD, one site is located on another chromosome), on their chromatin context (containing either active or repressive chromatin marks (Bonn et al., 2012) or located either in an A or a B compartment (Ogiyama et al., 2018) and that do not overlap any previously annotated transcript or regulatory region, to avoid unwanted deleterious effects of these insertions. Close-up genome browser plots of the different insertion sites are presented in Figures 20-21-22.

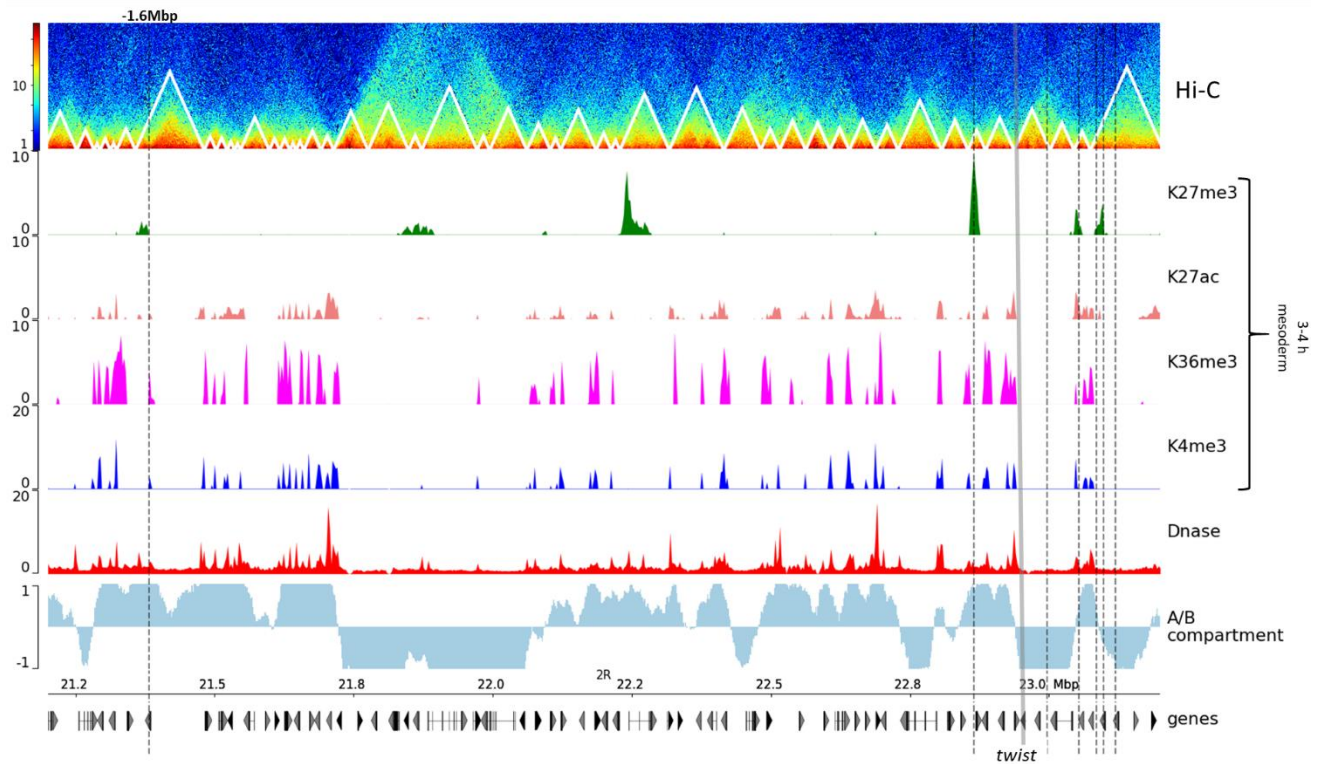
Distance to twist promoter	-1.6Mb	-181kb	Endogenous +3kb	+7.5kb	+39kb	+51kb	Chr3L
Distance to closest TAD border	678 bp	10.5 kb	4 kb	8.6 kb (same TAD)	10.2 kb (next TAD)	22.8 kb (next TAD)	22.7 kb
A (active)/B (repressive) compartment	A	A	A	A	A	B	A
Chromatin marks	Close to K27me3, K27ac, K4me3, K36me3	K27me3	K27me3 K27ac	K27me3 K27ac	Low K27me3	Close to K27me3	K27me3
DNase I hypersensitive region	No	No	Yes	No	No	No	No
Rescues the delta E3 lethality	NA	NA	NA	Yes	NA	Yes	No

**Table 2: Overview of the ectopic insertion sites of the E3 enhancer.**

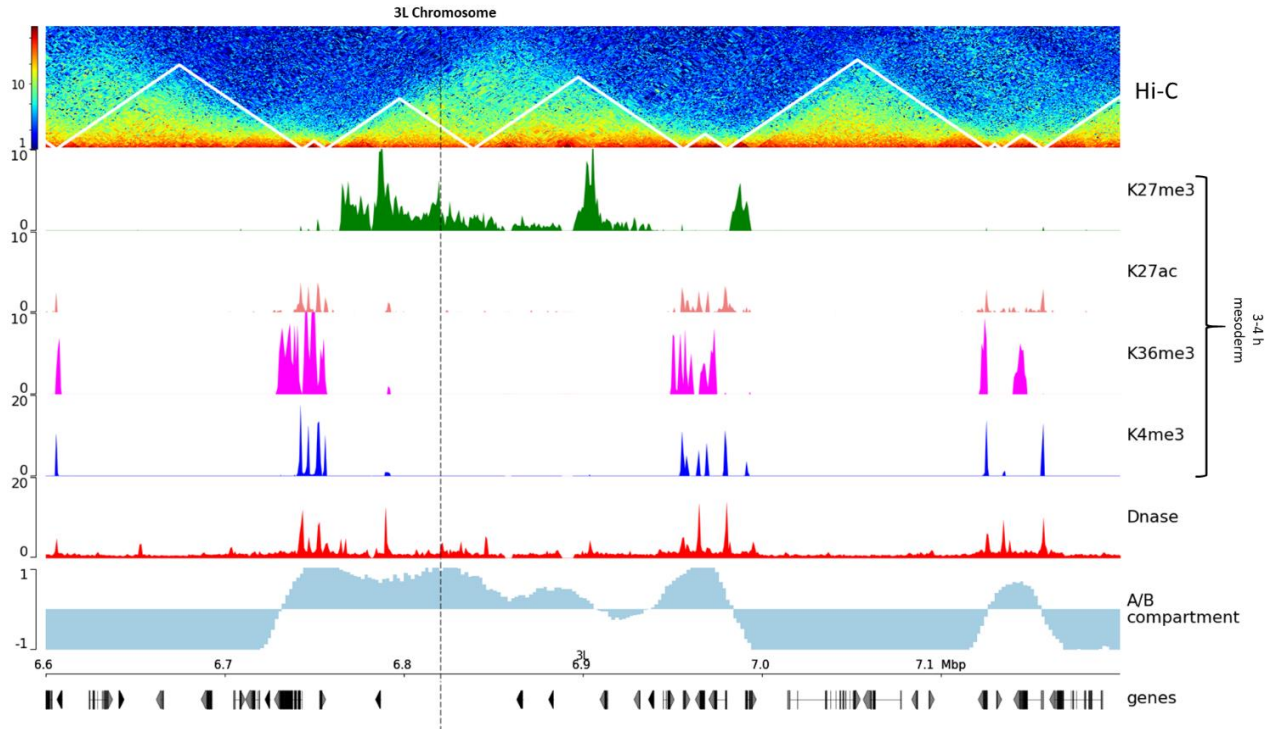


**Figure 20: Ectopic insertion sites of the E3 enhancer of twist on chromosome 2R, close to the twist locus.**

Genomic datasets plotted around the *twist* locus highlighting the diversity of transgenic locations. From top to bottom: Hi-C contact map in wild-type *Drosophila* embryos at stages 5 to 8 (Ogiyama et al., 2018). White triangle lines represent TAD domains defined by HiCEplorer (Ramírez et al., 2018).ChIP-seq chromatin marks from sorted mesodermal cells in 3 to 4 hours after egg-lay (Bonn et al., 2012). DNase-seq coverage in wild-type *Drosophila* embryos at stage 10 is shown in red (Thomas et al., 2011). A (Active) and B (Repressive) compartments (defined by HiCEplorer (Ramírez et al., 2018) using the Hi-C data from (Ogiyama et al., 2018)). Dashed lines: genomic location of the insertion sites. The distance between the insertion site and the *twist* promoter is indicated on top of the line. Continuous grey line: genomic location of the *twist* gene. Continuous light blue line: genomic position of the endogenous E3 enhancer.



**Figure 21: Ectopic insertion sites of the E3 enhancer of *twist* on chromosome 2R, far away from the *twist* locus.** Genomic datasets plotted around the *twist* locus highlighting the diversity of transgenic locations. From top to bottom: Hi-C contact map in wild-type *Drosophila* embryos at stages 5 to 8 (Ogiyama et al., 2018). White triangle lines represent TAD domains defined by HiCEplorer (Ramírez et al., 2018). ChIP-seq chromatin marks from sorted mesodermal cells in 3 to 4 hours after egg-lay (Bonn et al., 2012). DNase-seq coverage in wild-type *Drosophila* embryos at stage 10 is shown in red (Thomas et al., 2011). A (Active) and B (Repressive) compartments (defined by HiCEplorer (Ramírez et al., 2018) using the Hi-C data from (Ogiyama et al., 2018)). Dashed lines: genomic location of the insertion sites. The distance between the insertion site and the *twist* promoter is indicated on top of the line. Continuous grey line: genomic location of the *twist* gene.



**Figure 22: Ectopic insertion site of the E3 enhancer of twist on chromosome 3L.** The plotted genomic datasets highlight the diversity of transgenic locations. From top to bottom: Hi-C contact map in wild-type *Drosophila* embryos at stages 5 to 8 (Ogiyama et al., 2018). White triangle lines represent TAD domains defined by HiCExplorer (Ramírez et al., 2018). CHIP-seq chromatin marks from sorted mesodermal cells in 3 to 4 hours after egg-lay (Bonn et al., 2012). DNase-seq coverage in wild-type *Drosophila* embryos at stage 10 is shown in red (Thomas et al., 2011). A (Active) and B (Repressive) compartments (defined by HiCExplorer (Ramírez et al., 2018) using the Hi-C data from (Ogiyama et al., 2018)). Dashed line: genomic location of the insertion site.

## 5. The deletion of the endogenous E3 enhancer can be rescued by its ectopic re-insertion at given locations

As previously mentioned, the deletion of the endogenous E3 enhancer is homozygous lethal in *Drosophila* embryos. To test whether the insertion of E3 at the six ectopic sites could rescue this lethality, we deleted the endogenous E3 using CRISPR-Cas9 HDR and replaced it with a *3xP3::dsRed* cassette. Overall, we found that two out of three tested insertion sites could rescue the deletion of the endogenous E3 (Table 2): the insertion of E3 at + 7.5 kb from the *twist* promoter (this fly line will hereafter be called “+7.5kb\_DeIE3”), and at + 51 kb from the *twist* promoter (this fly line will hereafter be called “+51kb\_DeIE3”). While the rescue by the + 7.5 kb insertion site could be expected as it is located in the same TAD as *twist*, the rescue by the + 51 kb insertion site is much more surprising and unexpected. This indicates that the expression of *twist* can be regulated by an enhancer located in a different TAD at a distance that is 17 times greater than the original distance of the endogenous E3 enhancer to the promoter of *twist*. It should however be noted that the +51kb\_DeIE3 fly line is relatively weak and very sensitive to stress conditions. We are currently working on the finer characterization of this lines’ phenotype. Unfortunately, we have so far not been able to delete the endogenous E3 enhancer from the lines carrying an insertion site at position + 39 kb, - 181 kb, and - 1.6 Mb despite countless attempts and even sending the lines for injection to a private company. For the insertion site located on chromosome 3L, we simply crossed this line to the delta E3 line, and did not observe a rescue of the delta E3 lethality. Overall, this indicates that the E3 enhancer can regulate the expression of *twist* across large distances, even across a TAD boundary.

## **6. Enhancer-promoter interactions can occur across TAD boundaries and over large distances, even on a different chromosome**

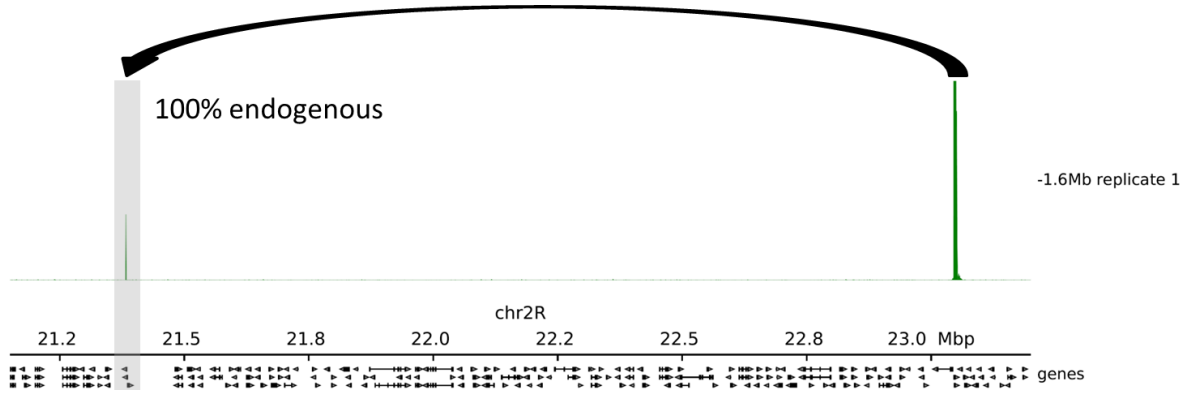
We next used the six fly lines in which the E3 enhancer was inserted at various distances from the *twist* promoter to assess whether these ectopically inserted enhancers could engage in long-range enhancer-promoter interactions with the *twist* promoter. For this purpose, we generated 4C-seq interaction maps from *Drosophila* embryos collected at 2 to 5 hours after egg-lay and used a viewpoint located just upstream of the *twist* promoter. As these fly lines contain two versions of the E3 enhancers, we created custom genomes that only contain one version of E3 located at the ectopic insertion site. We aligned the 4C-seq sequencing reads to this custom genome allowing mismatches, then performed variant calling to establish whether the reads overlapping a particular variant correspond to the endogenous or the ectopic version of the E3 enhancer. We then calculated the average percentage of reads corresponding to interactions with either the endogenous or the ectopic version of the E3 enhancer. In three out of six fly lines (corresponding to the insertions at -1.6Mb, -181kb, and +39kb to the *twist* promoter respectively; hereafter called “-1.6Mb”, “-181kb”, and “+39kb”), all the reads mapping to the E3 sequence contain the variants corresponding to the endogenous E3 sequence (Figure 23A-B-C). [Note: one of the replicates for line -1.6Mb needs to be repeated as the data was of poor quality]. This indicates that the *twist* promoter is not able to engage in long-range enhancer-promoter interactions with the E3 enhancer when it is inserted at these particular locations.

In the other three fly lines (corresponding to the insertions at +7.5kb and +51kb to the *twist* promoter and on chromosome 3L respectively; hereafter called “+7.5kb”, “+51kb”, and “chr3L”), at least a fraction of the reads mapping to the E3 sequence contain the variants corresponding to the ectopic E3 enhancer (Figure 24A-B-C). In the +7.5kb line, the ectopic E3 enhancer is located in the same TAD as *twist*, and as discussed above, this insertion can rescue the lethality of the endogenous E3 deletion. It is thus not very surprising to observe that the *twist* promoter can contact the E3 enhancer inserted at this location. More

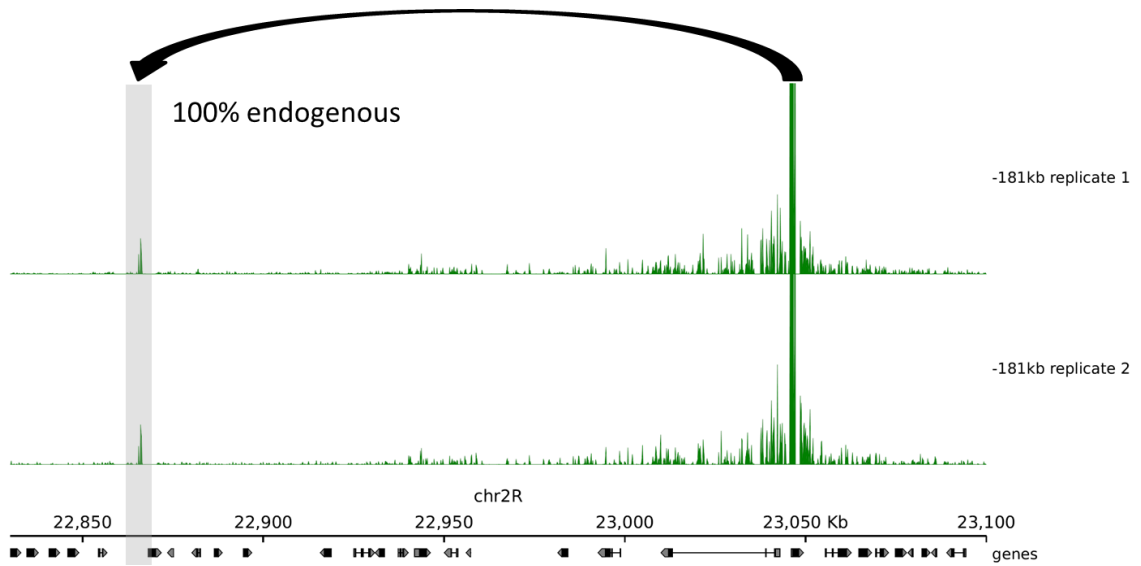
specifically, the *twist* promoter interacts both with the endogenous E3 enhancer (in 65% of the reads) and with the ectopic E3 enhancer (in 33% of the reads) (Figure 24A). In the +51kb line, the ectopic E3 enhancer is located in a neighboring TAD. However, despite the large distance and the presence of a TAD boundary in between, the *twist* promoter can engage in long-range interactions with the E3 enhancer when it is ectopically inserted at this location (Figure 24B). Interestingly, *twist* interacts with the E3 enhancer when it is located +51kb away, but not when it is located +39kb away, suggesting that distance alone cannot explain the formation of enhancer-promoter interactions. Even more surprising, 100% of the reads mapping to the E3 sequence correspond to variants present in the ectopically inserted E3 sequence, indicating that the *twist* promoter exclusively interacts with the ectopic E3 enhancer but no longer with the endogenous version. This interaction is in agreement with the rescue of the delta E3 lethality by the E3 ectopic insertion at position +51kb. More puzzling, the *twist* promoter can also interact with the E3 enhancer when it is located on a different chromosome, even though this *trans*-insertion is not sufficient to rescue the delta E3 lethality and hence regulate the expression of *twist*. In fact, in the chr3L line, the *twist* promoter interacts both with the endogenous E3 enhancer (in 51% of the reads) and with the ectopic E3 enhancer (49% of the reads) (Figure 24C).



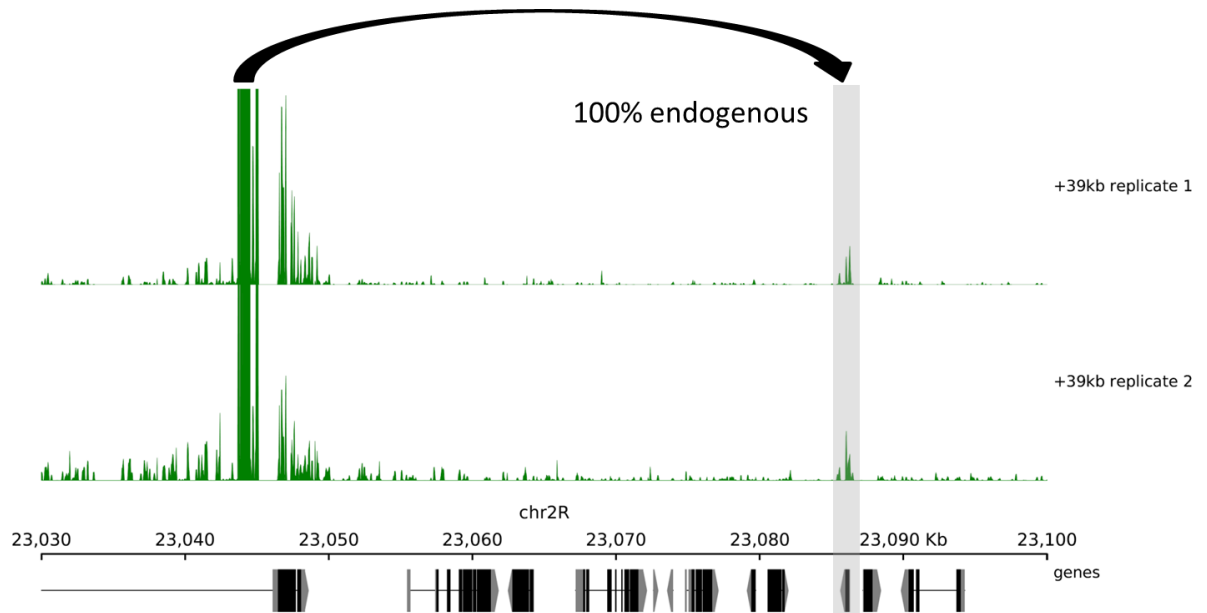
A



B

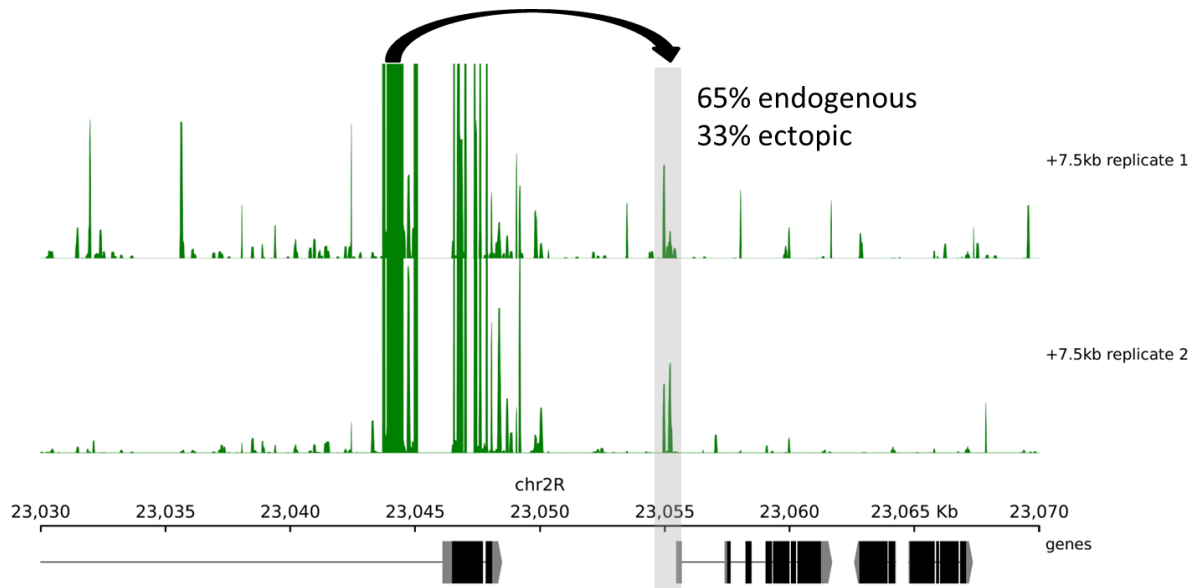


C

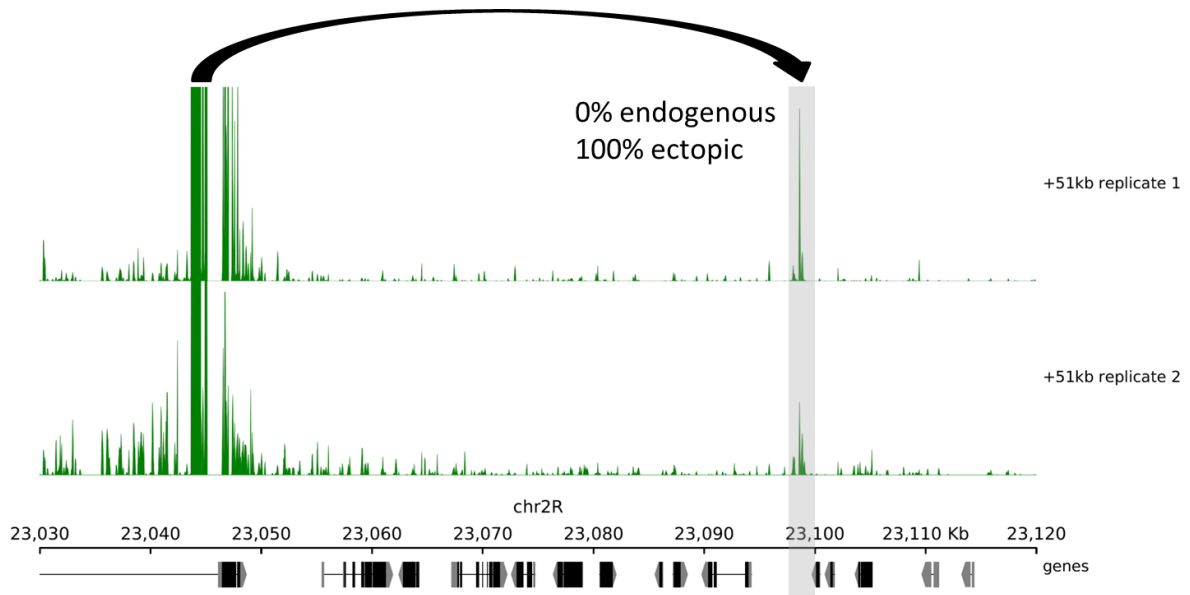


**Figure 23: The twist promoter doesn't interact with the ectopic E3 enhancer when inserted at locations -1.6Mb, -181kb, and +39kb.** 4C interaction maps in the -1.6Mb (A), -181kb (B), and +39kb (C) lines. *Drosophila* embryos were collected at 2 to 5 hours after egg-lay in two biological replicates. The observed interaction between the *twist* promoter (arrowtail; viewpoint) and the enhancer E3 (arrowhead) is highlighted in grey. The percentage of reads corresponding either to the endogenous or the ectopic version of the E3 enhancer is indicated.

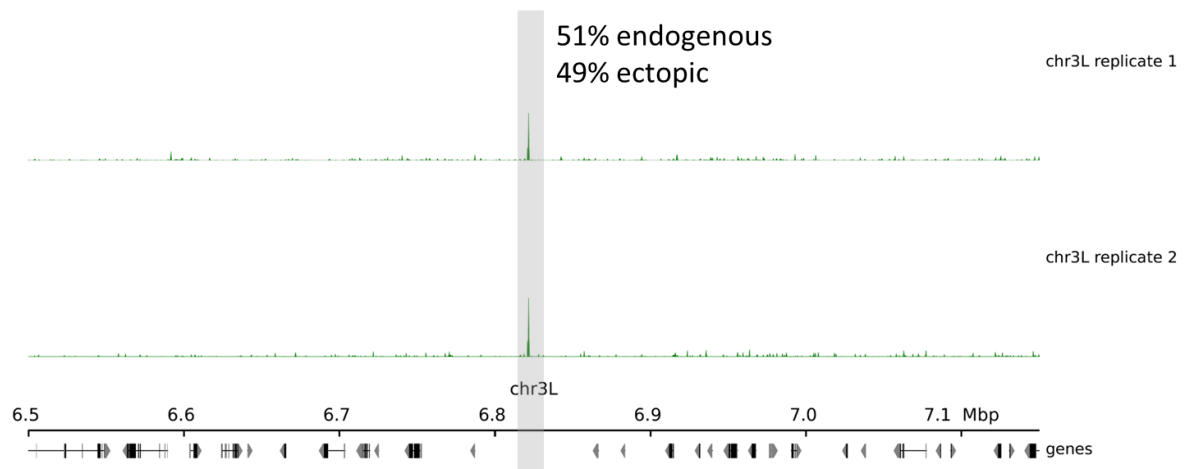
**A**



**B**



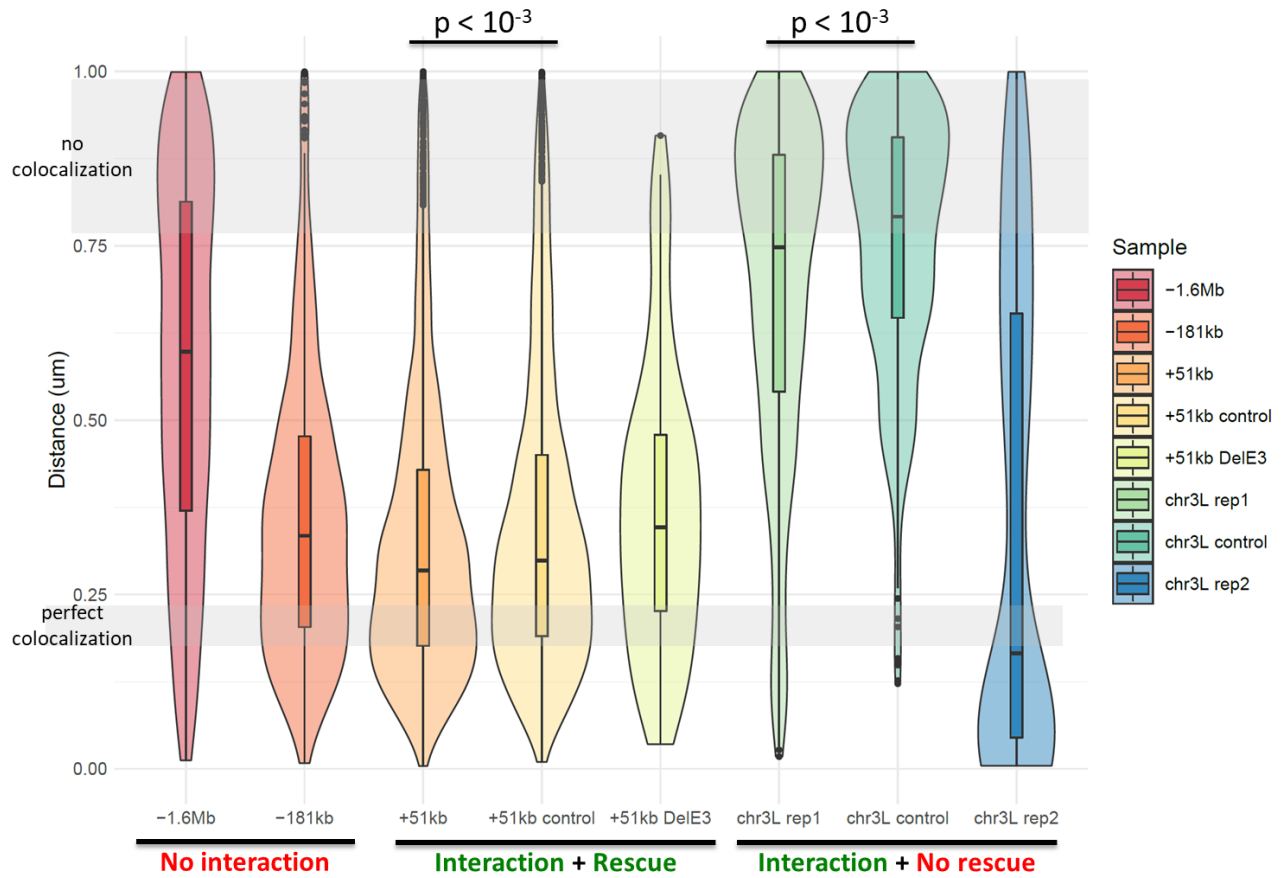
C



**Figure 24: The twist promoter interacts with the ectopic E3 enhancer when inserted at locations +7.5kb, +51kb, and on chromosome 3L.** 4C interaction maps in the +7.5kb (A), +51kb (B), and chr3L (C) lines. *Drosophila* embryos were collected at 2 to 5 hours after egg-lay in two biological replicates. The observed interaction between the *twist* promoter (arrowtail; viewpoint) and the enhancer E3 (arrowhead) is highlighted in grey. The percentage of reads corresponding either to the endogenous or the ectopic version of the E3 enhancer is indicated.

We further validated these interactions by 3D DNA fluorescence *in situ* hybridization (FISH) in *Drosophila* embryos (Figure 25). We used one probe located close to the *twist* promoter and one probe located close to the different insertion sites and measured the distance between the two probes across hundreds of nuclei. In the +51kb line, we observe that the distance between the two probes is significantly lower than the distance between the same probes in a control wild-type line ( $p < 10^{-3}$ ). This is in line with the 4C-seq results indicating that the *twist* promoter interacts with the E3 enhancer ectopically inserted at this position. In the chr3L line, we also observe that the distance between the two probes is significantly lower than the distance between the same probes in a control wild-type line ( $p < 10^{-3}$ ). However in this case, the average distances are much larger, as expected from their localization on two different chromosomes. Interestingly, we observe very high stochasticity in the distance between the two probes in the chr3L line. In one replicate (Figure 25, light green), a small proportion of cells exhibits a strong colocalization, while a larger proportion exhibits no colocalization. The opposite is true in a second replicate (Figure 25, blue). We are currently further investigating the basis of this stochasticity. As expected, the -1.6Mb and -181kb lines do not show a significant colocalization [Note: we are currently missing the controls for these two lines].

Together, these results indicate that an ectopically inserted enhancer can interact with its target gene in a distance-independent manner. This suggests that other molecular mechanisms could contribute to the specificity of enhancer-promoter interactions.



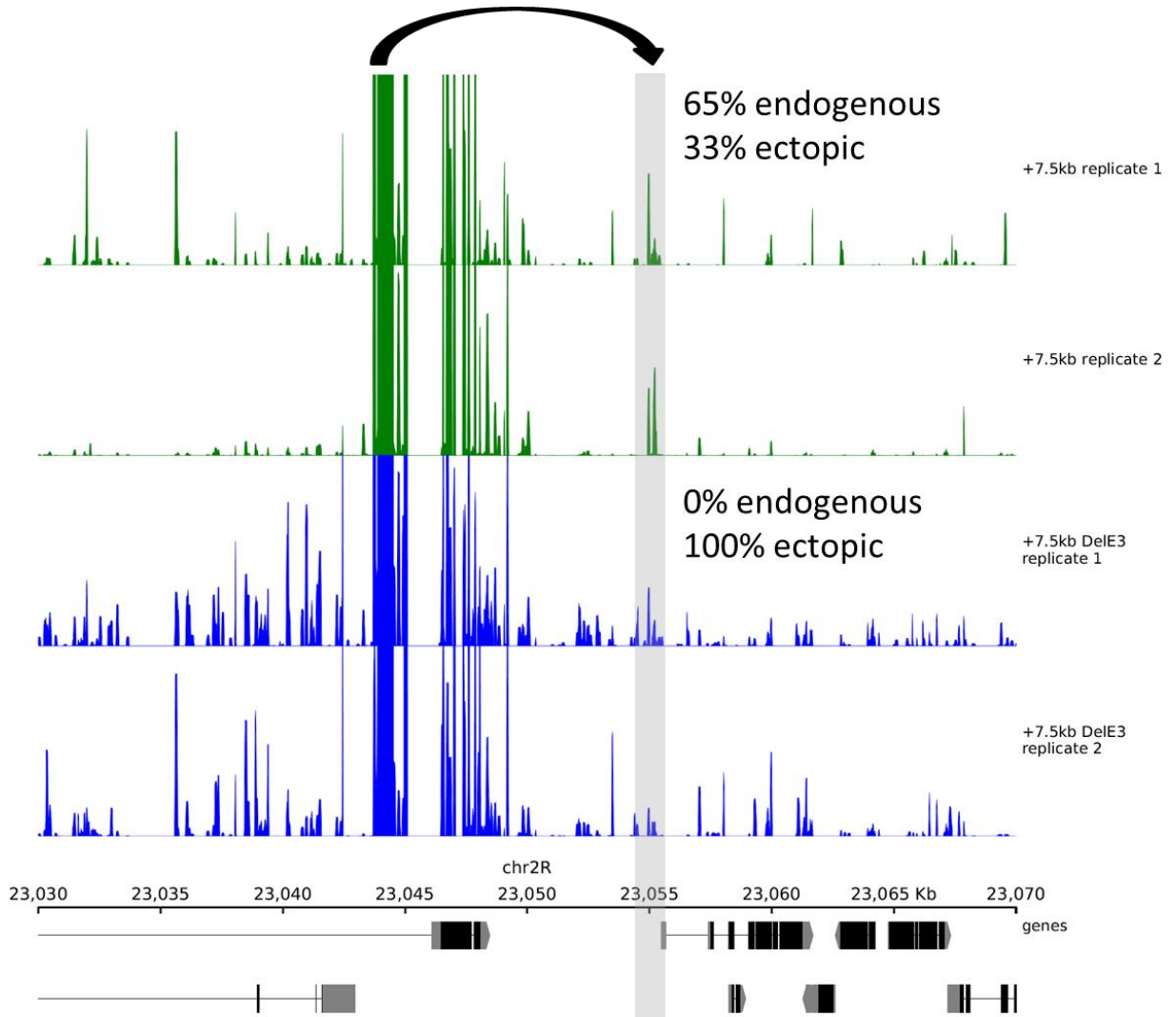
**Figure 25: Distance between twist promoter and the ectopically inserted E3 enhancer at different locations.** Violin plots indicating the distribution of inter-probes distances ( $\mu\text{m}$ ) between a probe located near the twist promoter and a probe located near the different insertion sites in lines -1.6Mb, -181kb, +51kb, +51kb\_DelE3, and chr3L. In the controls, the same probes were used in a wild-type line where the ectopic E3 enhancer is not present. The distances between the +51kb and chr3L lines and their respective control are significantly different ( $P < 10^{-3}$ , Wilcoxon test).

## 7. Long-range enhancer-promoter interactions are mediated by the presence of the endogenous E3 sequence

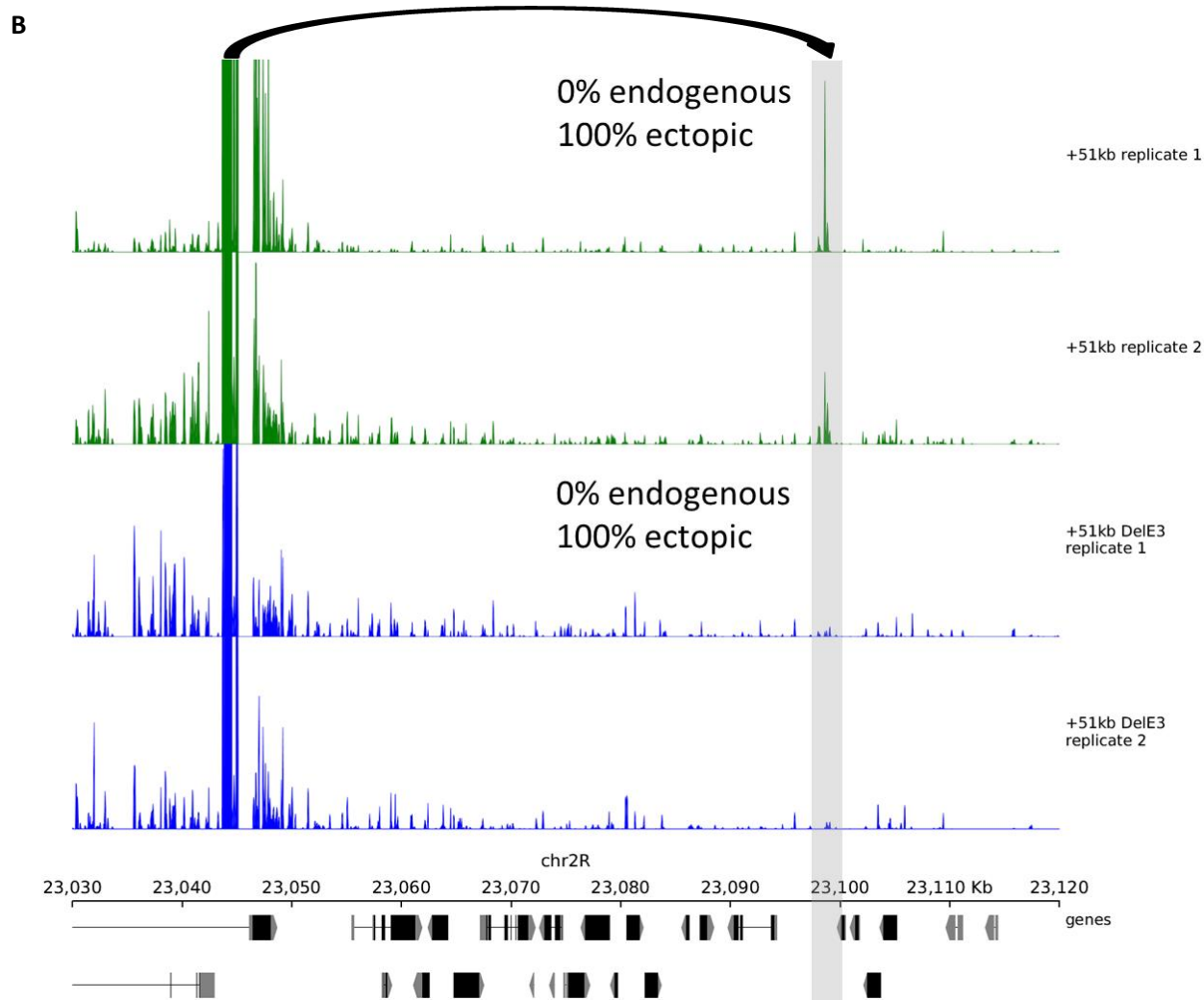
As mentioned previously, the insertion of the E3 enhancer at positions +7.5kb and +51kb can rescue the lethality of delta E3. We thus generated 4C-seq interaction maps from *Drosophila* embryos of the +7.5kb\_DeIE3 and +51kb\_DeIE3 fly lines collected at 2 to 5 hours after egg-lay and used the same viewpoint as previously (located upstream of the *twist* promoter). As previously, we mapped the 4C-seq sequencing reads to custom genomes where the endogenous E3 was deleted and used variant calling to assess whether the reads mapped to the endogenous or the ectopic version of E3. As expected given that the endogenous E3 sequence was deleted in those lines, 100% of the reads mapping in the E3 sequence correspond to the ectopic version of the E3 enhancer (Figure 26A-B). Surprisingly, however, the interaction between the *twist* promoter and the ectopic E3 enhancer is largely decreased compared to the lines where the endogenous E3 enhancer is still present. This observation was also confirmed by 3D DNA FISH in the +51kb\_DeIE3 line, where there is an increase in the distance between the probe located next to the *twist* promoter and the probe located next to the insertion site compared to the +51kb line (Figure 25). [Note: this observation needs to be further validated as it is based on only one embryo]. This decrease seems to be compensated by a local increase in background interactions around the *twist* locus (Figure 26A-B), but this preliminary observation requires further validation.

Overall, this indicates that the endogenous E3 sequence is required for the *twist* promoter to interact with ectopically-located enhancers. This is surprising as, in the +51kb line, the *twist* promoter was not even interacting with the endogenous E3 enhancer. However, even though the interaction is lost in these lines, the ectopically inserted E3 is still functionally sufficient to rescue the lethality of the endogenous E3 deletion.

A



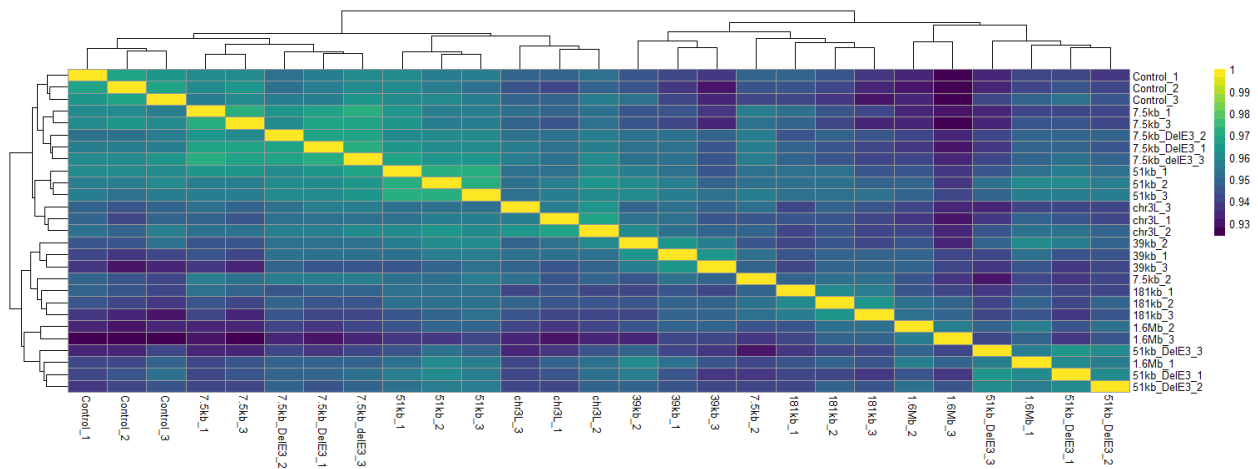




**Figure 26: The interaction between the twist promoter and the ectopic E3 enhancer is lost upon the deletion of the endogenous E3 enhancer.** 4C interaction maps in the +7.5kb (green) and +7.5kb\_DelE3 (blue) (A), and +51kb (green) and +51kb\_DelE3 (blue) (B) lines. *Drosophila* embryos were collected at 2 to 5 hours after egg-lay in two biological replicates. The expected interaction between the *twist* promoter (arrowtail; viewpoint) and the enhancer E3 (arrowhead) is highlighted in grey. The percentage of reads corresponding either to the endogenous or the ectopic version of the E3 enhancer is indicated.

## 8. The ectopic insertion of the E3 enhancer affects the expression of a large set of genes

The insertion of the E3 enhancer at ectopic locations might affect the expression of genes other than *twist*, via a mechanism called enhancer adoption or highjacking (Lettice et al., 2011; Northcott et al., 2014). Indeed, in some cases, an enhancer can misregulate the expression of nearby genes when it is inserted in a new regulatory landscape (Ghavi-Helm, 2019). To verify if such effect is present in the fly lines where we inserted the E3 enhancer in a different location, we performed RNA-seq on *Drosophila* embryos from all eight fly lines (i.e. +7.5kb, +7.5kb\_DelE3, +39kb, +51kb, +51kb\_DelE3, -181kb, -1.6Mb, chr3L) as well as in a wild type control, collected at 2 to 5 hours after egg-lay in three biological replicates (Figure 27). Differential expression analysis revealed that a large number of genes are either up- or down-regulated in either condition compared to the control (Table 3). Interestingly, the number of differentially-expressed genes (DEGs) seems to increase with the distance to the *twist* promoter, with, for example, the -1.6Mb and -181k lines having much more DEGs than the +7.5kb, +39kb, and +51kb lines. The lines where the ectopic E3 enhancer interacts with *twist* promoter (+7.5kb, +51kb, and chr3L) have fewer DEGs than the other lines, and this is even more accentuated when the ectopic E3 enhancer can rescue the lethality of delta E3 (lines +7.5kb and +51kb versus line chr3L). Finally, the deletion of the endogenous E3 in the +7.5kb, +51kb lines causes an increase in the number of DEGs, particularly in the +51kb\_DelE3 line with a seven-fold increase. This is in line with our observation that this fly line is relatively weak. This high number of DEGs in the +51kb\_DelE3 line can be explained by the fact that the *twist* gene itself is strongly down-regulated ( $\log_2$  fold change of -3.2), and as a consequence, all the genes that belong to the mesodermal gene regulatory network (Figure 28). Besides, the set of DEGs across all conditions highly overlaps (75.86 %) with a set of genes expressed in the mesoderm at 3 to 4 hours after egg-lay based on mesoderm-specific RNA-seq (Schor et al., 2018). This suggests that performing our RNA-seq experiments in whole embryos rather than in mesoderm-sorted cells did not significantly bias our analysis.



**Figure 27: Heatmap of the Spearman correlation coefficient for each RNA-seq library confirms the high quality of the data.**

	DEGs	DEGs with $ \log(\text{FC})  > 1$	% of upregulated genes	% of downregulated genes
<b>- 1.6 Mb</b>	4286	2221	37%	63%
<b>- 181 kb</b>	2645	1401	32%	68%
<b>+ 7.5 kb</b>	428	234	37%	63%
<b>+ 7.5 kb DeIE3</b>	733	317	35%	65%
<b>+ 39 kb</b>	1046	689	27%	73%
<b>+ 51 kb</b>	549	302	48%	52%
<b>+ 51 kb DeIE3</b>	4084	2028	35%	65%
<b>chr3L</b>	1156	649	35%	65%

**Table 3: Number of differentially-expressed genes in each condition.** Number differentially expressed genes (DEG) in each fly line ( $p < 0.001$ ) with or without a cutoff in the fold change (FC) of the differential expression, and percentage of up or down-regulated genes with a  $\log_2$  fold change  $> 1$  or  $< -1$ .



**Figure 28: Effect of the ectopic insertion on the expression of mesodermal genes.** Heatmap representing, for each condition, the log<sub>2</sub> fold change (FC) of differentially expressed known mesodermal gene (purple), mesodermal genes whose expression is down-regulated (blue) or not affected (green) in a *twist* mutant (based on data from (Sandmann *et al.*, 2007)).

We next assessed the location of the DEGs in each condition with respect to the location of the ectopic E3 insertion. Surprisingly, the DEGs seem to be spread out on the whole genome (Table 4), with no specific enrichment either in a +/- 50 kb window around the insertion site or on the same TAD as the insertion (Table 5). In fact, the closest DEG to the insertion site is located 3 to 10 kb away, or even 65 kb away in the case of the chr3L line. This suggests that the insertion of the E3 enhancer has widespread effects that stretch way beyond the TAD where this enhancer is inserted. One possible explanation for this high number of DEGs, is that the insertion of the ectopic 3 enhancer affects the expression of a small number of transcription factors, which will indirectly cause the misexpression of a large number of genes

from the same pathway. We thus performed GO term and KEGG pathway enrichment analysis, but this didn't reveal anything conclusive.

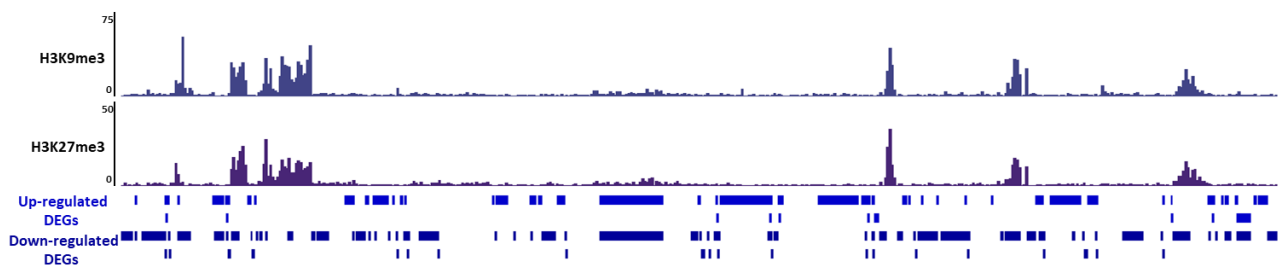
	-1.6Mb	-181kb	+7.5kb	+7.5kb DelE3	+39kb	+51kb	+51kb DelE3	chr3L
chrX	696	430	55	97	129	79	643	181
chr3R	1038	607	89	152	238	120	981	255
chr3L	868	557	47	141	230	129	874	265
chr2R	875	544	122	181	219	102	836	220
chr2L	778	479	113	160	223	116	717	228
Chr4	31	27	2	2	5	3	33	6
Total	4286	2644	428	733	1044	549	4084	1155

**Table 4: Number of differentially-expressed genes per chromosome in each condition.**

	+/- 50kb of the insert site			In the TAD of the insert			Closest DEG	
	Genes	DEGs	List of DEGs	Genes	DEGs	List of DEGs	Distance	Name
-1.6Mb	10	4	<i>dom, Acox57D-d, Acox57D-p, Sdc</i>	6	3	<i>Acox57D-d, Acox57D-p, Sdc</i>	4469 bp	<i>Acox57D-p</i>
-181kb	24	5	<i>stl, CG13531, fd59A, Gmer, nahoda</i>	18	4	<i>fd59A, Gmer, nahoda, Nup214</i>	3136 bp	<i>fd59A</i>
+7.5kb	19	2	<i>DCTN3-p24, PPO3</i>	8	2	<i>DCTN3-p24, PPO3</i>	8692 bp	<i>DCTN3-p24</i>
+7.5kb DelE3	19	1	<i>DCTN3-p24</i>	8	1	<i>DCTN3-p24</i>	8692 bp	<i>DCTN3-p24</i>
+39kb	24	4	<i>DCTN3-p24, l(2)k09913, CG3085, RpL37b</i>	24	2	<i>CG3085, RpL37b</i>	10121 bp	<i>CG3085</i>
+51kb	27	2	<i>CG9876, RpL37b</i>	24	2	<i>CG9876, RpL37b</i>	2942 bp	<i>RpL37b</i>
+51kb_DelE3	27	5	<i>twi, DCTN3-p24, CG43659, CG9876, RpL37b</i>	24	4	<i>CG43659, CG9876, RpL37b, side-V</i>	2942 bp	<i>RpL37b</i>
chr3L	3	0	-	3	0	-	63,987 bp	<i>lst1</i>

**Table 5: Differentially-expressed genes around the insertion site in each condition.** Total number of genes, number of differentially-expressed genes, and their name either in a +/- 50 kb window around the insertion site or on the same TAD as the insertion site. Distance and name of the closest differentially-expressed gene to the insertion site.

We then attempted to identify any common characteristics of our sets of DEGs. For this purpose, we examined the overlap between the location of up- and down-regulated genes and the presence of activating or repressive chromatin marks. Up-regulated genes are highly depleted from regions that exhibit repressive chromatin marks (an example is shown in Figure 29), suggesting the presence of the E3 enhancer is not sufficient to activate the expression of a gene that was actively repressed.



**Figure 29: Up-regulated genes are depleted from regions enriched in repressive chromatin marks.**

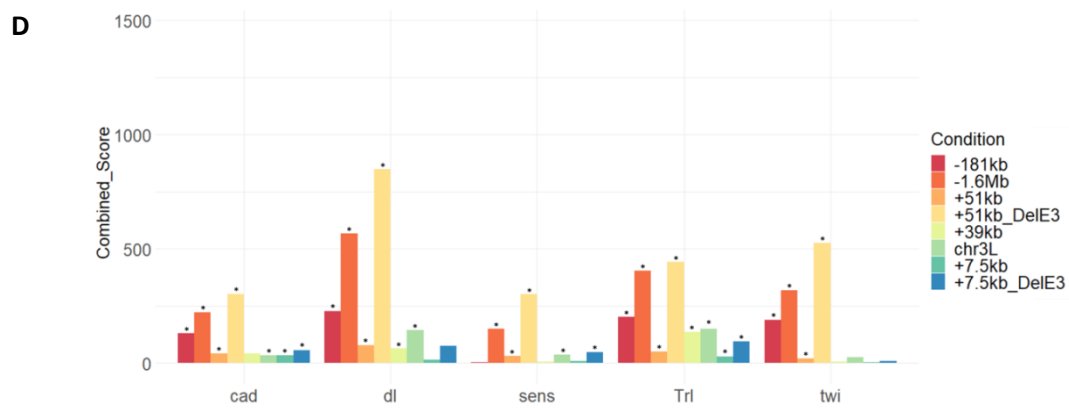
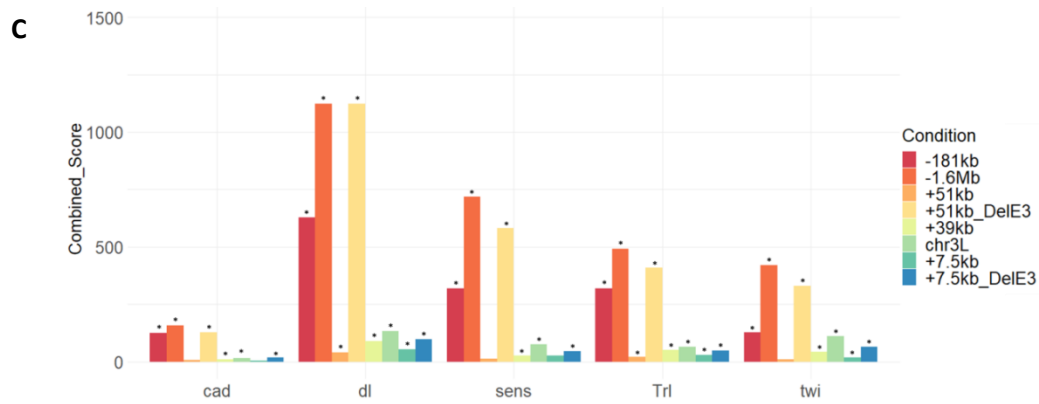
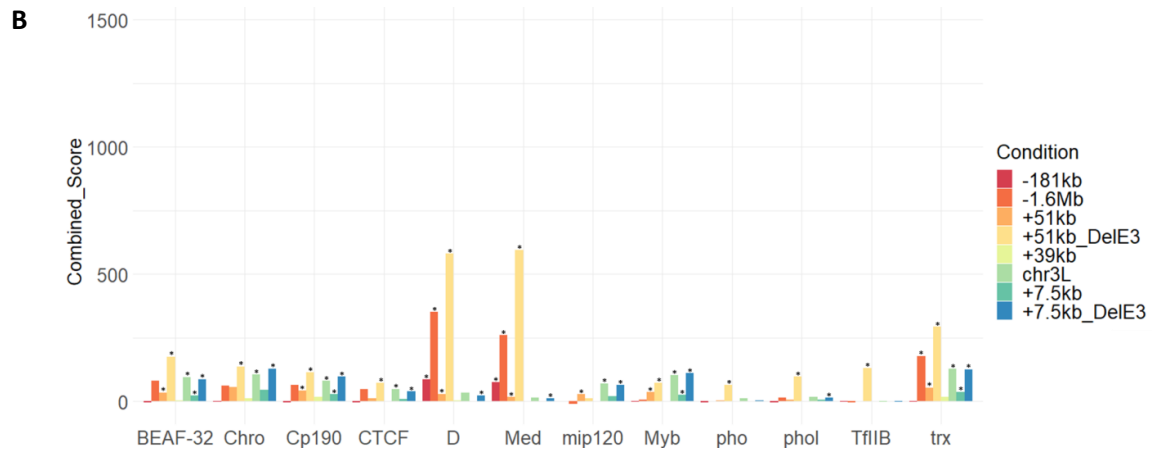
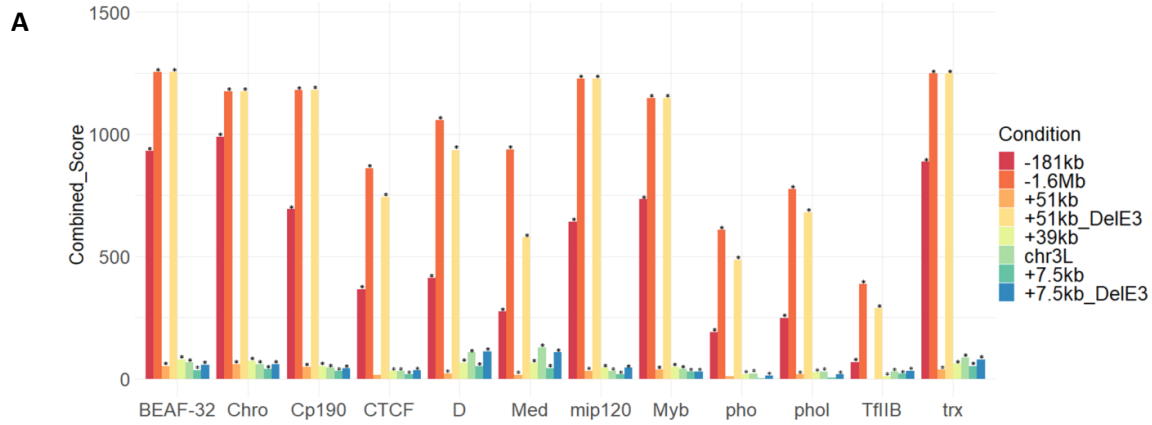
Example genome browser plot showing the overlap between the H3K9me3 and H3K27me3 repressive chromatin marks (Nègre et al., 2011) and the location of up- (light blue) and down- (dark blue) regulated genes across all conditions.

Finally, we used the FlyEnrichr database (Chen et al., 2013; Kuleshov et al., 2016), to examine if up- and down-regulated genes are enriched for the binding of any specific transcription factor (Figure 30). Overall, up-regulated genes are significantly enriched for the binding of several transcription factors including the BEAF-32, Chromator (Chro), Cp190, and CTCF insulator proteins Figure 30 A versus 30 B). On the other hand, the binding of other transcription factors, such as caudal (cad), dorsal (dl), senseless (sens), Trithorax-like (Trl), and twist (twi) is enriched both at up- and down-regulated genes (Figure 30 C-D). This suggests that the genes that can be directly or indirectly positively regulated by the E3 enhancer are bound

by a given set of insulator proteins. At this point, this observation is only a correlation, which will require further investigations. We are for example planning to complement this differential expression analysis with 4C-seq or Micro-C interaction maps focused on the ectopically inserted E3 enhancer.

## **9. Conclusion**

Overall, our results show that: (i) The E3 enhancer functions in an orientation-independent manner at its endogenous locus. (ii) The *twist* promoter can interact with the E3 enhancer, even when it is ectopically inserted at distant locations in a different TAD or even in a different chromosome. These ectopic insertions can sometimes rescue the lethality caused by the deletion of the endogenous E3 enhancer. Surprisingly, the interaction with the ectopic E3 enhancer is lost upon the deletion of the endogenous version, with no effect on the rescue. (iii) The ectopically-inserted E3 enhancer can misregulate the expression of a large number of genes, with no specific enrichment around the insertion site. Why only certain insertion sites can interact with the *twist* promoter remains an open question which we are currently investigating.





**Figure 30: The binding of specific transcription factors is enriched at up- and down-regulated genes.**

Combined score obtained from Fly Enrichr (based on the p-value obtained by Fisher's exact test and the z-score) (Chen et al., 2013; Kuleshov et al., 2016) for two sets of transcription factors whose binding is enriched at up-(**A-C**) and down-(**B-D**) regulated genes. The asterisk highlights conditions where the enrichment of a particular transcription factor binding is significant compared to a similarly-sized set of randomly-sampled genes.



# Chapter 3: Discussion



During the development and differentiation of an organism, precise gene regulation is essential for cells to maintain and control their differentiation processes. This complex regulation is mediated by *cis*-acting DNA elements, which include enhancers and promoters.

**1. Inverting the orientation of an enhancer at the endogenous locus does not influence the expression of the target gene.**

Studies mostly based on enhancer-reporter assays have revealed that enhancers can function independently to their orientation and location. In these experiments, an enhancer cloned upstream of a reporter gene in both orientation and different locations is still able to recapitulate the expression of the reporter without the formation of a chromatin loop (Banerji et al., 1981). Whether the orientation of an enhancer could affect the expression level of its target gene in a native genomic context remains however elusive.

When the endogenous enhancer E3 orientation has been inverted, it did not cause any lethality in embryogenesis and did not affect *twist* expression. This supports the notion that even at original endogenous genomic loci, enhancers can function in an orientation-independent manner. Furthermore, the data showed that the inverted orientation of E3 enhancer did not affect either the chromatin interactions. However, as the distance (around 3kb) between the endogenous E3 enhancer and *twist* promoter makes it difficult to detect accurate chromatin interactions due to the limitations of 4C technology, it would be interesting to validate the data further by performing the new high-resolution technology Micro-Capture-C. This new technology allows us to determine the physical chromatin contacts between regulatory elements at a resolution of individual transcription factors, providing new information in gene regulation (Hua et al., 2021). Thus, discoveries of higher resolution assays may confer new insights on enhancer-promoter interactions.

**2. Enhancer and promoter can interact across boundaries and sometimes can rescue the phenotype upon the deletion of the essential regulatory element.**

An aspect that is still under investigation is whether a given enhancer located at different genomic locations could still form a chromatin loop with its target promoter. More precisely, understanding how enhancers choose to interact with target promoters has been challenging because they can be located at large genomic distances along chromosomes.

As it is mentioned before, one model of how a distal enhancer can activate a gene target is through looping. Proteins that facilitate this contact such as the mediator complex propagate the looping and the physical interaction between enhancers and promoters. Together, promoter and enhancers constitute the regulatory landscapes of genes and their looping seems to be controlled by the spatial configuration of chromatin into TADs (Andrey and Mundlos, 2017).

Genome architecture studies, genome engineering, and imaging approaches have contributed to extraordinary progress in understanding the molecular mechanisms of enhancer-promoter interactions` specificity in the context of TAD domains (Kim and Dean, 2021). The effects of TAD domains on enhancer-promoter communication can be considered essential for understanding the relationship between chromatin topology and gene regulation.

However, there are currently divergent hypotheses on whether TAD domains are required for proper enhancer-promoter communications or whether they do not restrict enhancer-promoter communication (Laugsch et al., 2019a).

Previously, it has been shown that chromatin interactions between promoters and enhancers are mostly constrained within an entire TAD domain rather than neighboring TADs. There is considerable evidence that TAD boundaries act as insulators to induce proper enhancer-promoter interactions while precluding inappropriate ones. Indeed, an enhancer that is located within a TAD prefers to interact with promoters

of the same TAD rather than with those located in neighboring TADs (Symmons et al., 2014b). In agreement with this, genes located within the same TAD domain are expressed coordinately and alterations in TAD boundaries can cause the formation of new enhancer-promoter interactions that alter gene expression (Stadhouders et al., 2019). For example, deletions or inversions of TAD boundaries can lead to their repositioning and/or relocation of enhancer elements into other compartments, affecting gene expression by impairing enhancer-promoter interactions and often by causing developmental abnormalities and disease (Spielmann et al., 2018). Several studies showed that transcription is abolished when a TAD boundary is located between enhancer and promoter. This indicates that TAD boundaries can insulate and reduce the chromatin interactions frequency between enhancers and promoters (Andrey and Mundlos, 2017). Recently it has been demonstrated that the transcription level of the gene decreases with the increased genomic distance between the enhancer and the promoter from the TAD domain until reaching the outside of the TAD boundary. Furthermore, also contact probabilities sharply decayed with an increasing genomic distance of the misplaced enhancer from the promoter until approaching the TAD boundary and further fell drastically across boundaries. Together these data show that the activity of the enhancer is delimited by TAD boundaries (Zuin et al., 2021).

On the other hand, perturbation of insulator elements at TAD boundaries has little effect on genome-wide gene expression, indicating that enhancer-promoter interactions are independent of TADs. In support of this, placing an enhancer in a different TAD domain, through inversion of TAD structure is not sufficient to form new enhancer-promoter interactions (Laugsch et al., 2019a;). Similarly, in *Drosophila*, balancer chromosomes, which contain many inversions, deletions, and mutations, only mildly affect gene expression during embryonic development (Ghavi-Helm et al., 2019). This indicates that the recruitment of a new gene target by an enhancer following genomic rearrangements is highly locus-dependent. Another recent example where TAD boundaries do not prevent communication between cis-regulatory

elements occurs for example in mammals with the *xist* locus, in which its regulatory landscape is not restricted to its own TAD domain but includes cis-regulatory elements located in the neighboring TAD (Galupa et al., 2020).

All these controversial results suggest that TADs can play an important role in enhancer-promoter communication, however, other molecular mechanisms might be involved in enhancer-promoter interaction specificity.

After generating transgenic fly lines where the endogenous enhancer E3 is ectopically inserted at various distances/locations from the *twist* promoter. We observed that out of six transgenic fly lines, only three show chromatin interactions (Table 6). **1)** More specifically, in the fly line where the E3 enhancer is located at 7.5 kb distance from its promoter, the 4C data show chromatin interactions between *twist* promoter and both at the endogenous enhancer E3 (65%) and at the ectopic enhancer E3 (33%). We reasoned that the interaction with the ectopic enhancer E3 is due to its location within the same *twist* TAD domain. This confirms previous studies that demonstrated that TAD domains facilitate the interactions between cis-regulatory elements (Andrey and Mundlos, 2017; Symmons et al., 2016; Zuin et al., 2021). Nevertheless, chromatin interactions observed for the other two fly lines do not support this first thought. **2)** Indeed, the second transgenic fly line that shows a chromatin interaction with the new ectopic E3 enhancer located +51 kb far from *twist* promoter is located in a different TAD domain. In this case, the chromatin interaction occurs 100% ectopically skipping the endogenous enhancer E3. This suggests that the promoter of *twist* prefers to interact with a distal ectopic enhancer instead of the closest one. Thus, our results are in line with the previously published data that even in the compact genome of *Drosophila*, around 30% of enhancers function over large genomic distances (10-100kb) (Furlong and Levine, 2018). **3)** The third transgenic fly line where the ectopic E3 enhancer is located on a different chromosome (3L), *twist* promoter interacts both with the endogenous enhancer (51%) and with the ectopic one (49%). This



*trans*-interaction confirms that *twist* promoter can interact with the ectopic enhancer E3 located on a different chromosome.

In the other three fly lines (1.6Mb, -181kb, +39kb) *twist* promoter is not able to engage in long-range enhancer-promoter interactions with the ectopic E3 enhancer. In the first place, we thought that the presence of the chromatin mark H3K27me3 with polycomb proteins might influence the interaction to happen. However, H3K27me3 is also present in the fly lines where the interaction is occurring. We think that other factors located near the ectopic insertion are enabling the interactions to occur. For example, one feature that these three fly lines have in common is that they are close to a TAD boundary. Taking into consideration that architectural proteins, insulators, and highly transcribed genes are mainly enriched at TAD boundaries. One possible explanation is that insulator elements may promote the ectopic E3 enhancer to interact with other non-native promoter genes.

Overall, our results identified long-range chromatin interactions between *twist* and ectopic E3 enhancer across TAD boundaries and along the genome. This reinforces the above theories that TAD boundaries do not necessarily restrict enhancer-promoter interactions but other mechanisms can be considered as a driver (Calhoun and Levine, 2003; Schoenfelder and Fraser, 2019). Therefore, a current central challenge in genome biology is to identify the principles and the mechanisms that cause enhancer-promoter specificity. In particular, how do promoters choose to interact with a distal enhancer ignoring the most proximal ones?

Taking into consideration that the deletion of endogenous E3 enhancer is embryonic lethal, what surprised us is that this deletion is affecting the newly established chromatin interactions. Indeed, in the fly line (+51kbDelE3) the newly established interaction is decreased with at the same time an increase of background interactions, which is sufficient to rescue the phenotype. This indicates that other molecular mechanisms could play a role in stabilizing the looping. For example, we reasoned that the endogenous

E3 enhancer is facilitating the interaction frequency between *twist* promoter and the ectopic E3 enhancer, even if it is not required for the interaction to occur. In the +7.5kbDelE3fly line, upon the deletion of the endogenous E3 enhancer, *twist* forms a ‘complete’ chromatin interaction (100%) with the ectopic E3 enhancer, this interaction is decreasing but still sufficient to rescue the phenotype. However, it is important to note that as the 4C technology is not a quantitative method, it would be interesting to validate further with higher resolution assays.

In the chr3L fly line, *trans*-interaction between *twist* promoter (located at chromosome 2R) and the ectopic E3 enhancer (located at chromosome 3L) was not sufficient to rescue the phenotype. One possible explanation is partially showed by 3D DNA FISH data where two biological replicates, where we observed two degrees of significant colocalization. The number of nuclei that confirmed the interaction is variable between replicates, meaning that in one case a high number of nuclei confirmed the interactions compared to the other case where the number of nuclei showing the interaction was lower. Therefore, it is necessary to repeat the experiments by introducing more replicates; this will help us to fully validate the 4C data.

	Fly lines where the ectopic enhancer E3 is interacting with <i>Twist</i> promoter				
	+7.5Kb	+7.5KbDelE3	+51kb	+51kbDelE3	Chr 3L
Chromatin interaction with endogenous enhancer E3	65%				51%
Chromatin interaction with ectopic enhancer E3	33%	100%	100%	100%	49%
Rescue of the phenotype		YES		YES	NO
Decay of previously established chromatin interaction		YES		YES	

**Table 6: Features of fly lines where the ectopic E3 enhancer is interacting with *Twist* promoter.** The percentage indicates the degree of chromatin interaction. Del (deletion of the endogenous enhancer E3). +kb (distance from the insertion site to the *twist* promoter). % (percentage of reads).

This variability between all these three transgenic fly lines, which are shown in table 6, led many questions on how the ectopic E3 enhancer can interact with the *twist* promoter. First, we hypothesized that transcription factories may play a role in enhancer-promoter interactions. Indeed increasing the concentration level of different transcription factors, which sub-sequentially form machinery hubs, can promote the interaction between a distal enhancer and the promoter. In most cases, DNA loop formation between enhancers and core promoters occurs as a consequence of the interaction between enhancer-bound transcription activator, mediator, and promoter-bound RNA polymerase II (Furlong and Levine, 2018; Kagey et al., 2010). This can explain the stability of the loops in conditions with both the presence and the absence of the endogenous E3 enhancer.

Regarding the chromatin interactions that occur across TAD boundaries in our different conditions, we thought that insulator elements located close to the insertion sites could play a role in the formation of long-range enhancer-promoter interactions. Recently with the advent of new technology such as Micro-C and single-cell Hi-C, it has been argued that TAD and their boundaries are variable from cell to cell. This may allow the enhancers to skip the insulator elements located at boundaries and contact regions outside of a TAD (Hsieh et al., 2021). Indeed, the study argued that long-range enhancer-promoter interactions across TAD boundaries, was a consequence of the level of CTCF or cohesin occupancy, which if it was lower it shows a weaker insulation activity. However, they demonstrated that the level of CTCF and RAD21 occupancy was almost the same in both scenarios where enhancer-promoter contacts were occurring across or not across TAD boundaries. They conclude that both TAD-dependent and TAD-independent mechanisms contribute to the regulation of enhancer-promoter interactions (Hsieh et al., 2021).

The local chromatin structure, which includes accessibility, histone marks, transcription binding of enhancer and candidate target promoters, might also play important role in defining the specificity of these interactions.

Despite previous studies, which showed that relocation of enhancers through TAD rearrangements does not have a mild effect on gene expression. Our results demonstrated that in all transgenic fly lines generated, the insertion of E3 enhancer has a huge impact on differentially expressed genes across the entire genome. This can be a consequence of regulatory elements, which directly and/or indirectly affect other important transcription factors pathways influencing the expression of genes. However, even if the majority of genes are differentially expressed, the fly lines are still surprisingly viable. Indeed, it would be necessary to investigate further, what might be the cause of the high number of differentially expressed genes across the genome. Moreover, in a normal scenario, *twist* expression is regulated by the three enhancers, however, in the +51kbDelE3 fly line we observed from the RNAseq data the downregulation of *twist*. This can be explained by the fact that the newly established interaction is only enough to rescue the phenotype, but is not sufficient to bring up the normal transcription level of *twist*. We think that our results follow the model where looping is necessary but not sufficient for transcription. This is not in line with other studies where the lifetime of enhancer-promoter colocalization increases when transcription is activated, supporting the model that a high local concentration of transcription factors can induce transcription burst facilitating enhancer-promoter communication (Field and Adelman, 2020).

To recapitulate, our results indicate that in *Drosophila melanogaster*, enhancers function in an orientation-independent manner, even at their endogenous locus. Surprisingly, enhancer-promoter interactions can sometimes occur across TAD boundaries and even with an enhancer located on different chromosomes. Finally, we observed that the ectopic insertion of an enhancer has a huge impact on gene expression, causing a large number of differentially expressed genes along the genome.

Taken together these results with the integration of additional experiments will provide a better understanding of enhancer biology, chromatin organization, transcription regulation, and enhancer-promoter interaction specificity during early *Drosophila* embryogenesis.

### **Perspectives**

Following the results of this thesis, future directions are needed to investigate deeper the molecular mechanisms that drive enhancer-promoter interaction specificity and to provide unique insights on E3 enhancer activity. With these data, we will be able to clarify the role of the genomic context in the formation of enhancer-promoter interactions in early *Drosophila* embryogenesis.

Additional computational analysis and experiments are essential to identify transcription factor binding sites, insulator-binding sites, and chromatin marks in the vicinity of the enhancer E3 insertion sites. These different factors can potentially mediate *twist* promoter and ectopic enhancer interactions as well as could influence the chromatin organization. For example, performing chromosome conformation assay where the viewpoint is the ectopic E3 enhancer will help us to identify new interaction with non-native promoters, which at the same time could explain specific differential expressed genes.



# Chapter 4: Materials & Methods





## 1. Material

### 1.1 Equipment

Product	Supplier
Eppendorf centrifuge	Eppendorf
Thermomixer	Eppendorf
Water Bath	Thermo Fisher
50 NucleoMag SEP Maxi magnetic separation Rack	Macherey Nagel
12-Tube Magnetic Separation Rack	New England Biolabs
Qubit HS/BR	Invitrogen
T100 Thermal Cycler	BioRad
TapeStation	Agilent
FemtoJet 4i electronic microinjector	Eppendorf
Model P-1000 Micropipette Puller	Sutter Instrument

**Table 7: List of equipment**

### 1.2 Plasmids

#### *Plasmids used in this study*

Name	Supplier
<i>pJET1.2-STOP-dsRed</i>	Addgene #60944 (Schnorrer lab)
<i>pHD-dsRed</i>	DGRC #1360 (O'Connor-Giles lab)
<i>pU6-BbsI-gRNA</i>	Addgene #45946 (O'Connor-Giles lab)
<i>pHD-DsRed-attP</i>	DGCR #1361 (O'Connor-Giles lab)
<i>pBS-KS-attB1-2-PT-SA-SD-0-2xTY1-V5</i>	Addgene #61255 (Schnorrer lab)
<i>p3xP3-EGFP.vas-int.NLS</i>	Addgene #60948 (Schnorrer lab)

<i>pattB</i>	DGRC # 1420 (Bischof and Basler lab)
--------------	--------------------------------------

**Table 8: List of plasmids used in this study**

*Plasmids generated during this study*

<i>pHD-dsRed-2attP</i>	Backbone of the donor vector for the deletion of E3
<i>pHD-dsRed-2attP-HA1&amp;2</i>	Donor vector for the deletion of E3 with homology arms
<i>pU6-BbsI-gRNA5'</i>	gRNA vector recognizing the PAM site located 5' from E3
<i>pU6-BbsI-gRNA3'</i>	gRNA vector recognizing the PAM site located 3' from E3
<i>pBS-KS-attB-E3</i>	<i>attB</i> vector for the re-insertion of E3 in either orientation
<i>pHD-dsRed-attP-HA1&amp;2</i>	Donor vector for the insertion of an <i>attP</i> site + 7.5kb away from the <i>twist</i> promoter, with homology arms
<i>pU6-BbsI-gRNA</i>	gRNA vector for the insertion of an <i>attP</i> site + 7.5kb away from the <i>twist</i> promoter
<i>pattB-E3</i>	<i>attB</i> vector for the insertion of E3 + 7.5kb away from the <i>twist</i> promoter
<i>pBS-KS-attB-E3_MIMIC</i>	<i>attB</i> vector for the insertion of E3 in the MIMIC lines

**Table 9: List of plasmids generated during this study**

**1.3 Primers**

<b>Primers used for the deletion of the endogenous enhancer E3</b>	
SV788	CGTACGCCAGGTCAGAAGCGGTTTTTC
SV789	CGTACGAACCCCTTGTGTCATGTCGG
SV790	ACCGGTAACCCCTTGTGTCATGTCGG
SV792	ACTAGTCCCAGGTCAGAAGCGGTTTTTC
SV797	NNNNCACCTGCNNNNTCGCTGAGTTATCTGTTCCGGGGTGT

SV798	NNNNCACCTGCNNNNTTATTACAGGCGATGTGTGGAGCAA
FW799	ACTAGTCAGAGGGTTGCAAGTGCATTAGGC
RV800	CTGCAGGGAATCGGATCCATTCCGGTATGG
gRNA_5'E3	GAAATCAAAGACTTGTATACAGG
gRNA_3'E3	GGGTAAAAATATCTTTGCAGAGG
<b>Primers used to amplify the endogenous enhancer E3 (to generate the <i>pBS-KS-attB-E3</i> vector)</b>	
SV908	gtccatcatgatggctgactctagaTACAAGTCTTTGATTCTG
SV858	GTCCATCATGATGGTCGACAAGCTTCAAAGATATTTTTACCCCCATAACAGAG
<b>Primers used to add an <i>attP</i> site +7.5kb away from the <i>twist</i> promoter</b>	
HA1_FW	ATTCCACCTGCATTCTCGCAGATCGTATCGCGCATTCT
HA1_RV	ATTCCACCTGCATTCTACAAGATATGCCCGACATTCGAT
HA2_FW	ATTCGCTGTTTCATATTGGAGCCTTGTGCTATTGGT
HA2_RV	ATTCGCTCTTCAGACCATTCTTTGGCTCCGATCG
gRNA_attP	TCGAATGTCGGGCATATCTTTGG
<b>Primers used for the amplification of endogenous enhancer E3 (to generate the <i>pattB-E3</i> vector)</b>	
FW_KpnI	ATTCggtaccATGTTTTTCCAACCTTTGTTTGTCC
RV_XhoI	ATTCctcgagCAAAACAAAAAAGTGGGGAAATTTCCG
<b>Primers used for the amplification of endogenous enhancer E3 (to generate the <i>pBS-KS-attB-E3_MIMIC</i> vector)</b>	
loxp_28FW_E3_Alexia:	tctagaNNNNATAACTTCGTATAATGTATGCTATACGAAGTTATNNNNATGTTTTTCCAAC TCTTGTTTGTCC
E3_HindIII_R	ATTCaagcttCAAAACAAAAAAGTGGGGAAATTTCCG
<b>3D DNA FISH probes</b>	
DNAFISH_twi1_FW	TGTGAGCAGTTGTTTACCGC
DNAFISH_twi1_RV	GAGGGGCATTGTACGACTCT
DNAFISH_twi2_FW	TTTAAAGCGTCGCACCAACA
DNAFISH_twi2_RV	GGCAAACCTCCACAGCTTTT
DNAFISH_twi3_FW	TGCAGCGTGTCCAATTTGAA

DNAFISH_twi3_RV	CTGGGAAGTATGCGTGTGG
DNAFISH_twi4_FW	CAACATACCCAGCAACGAGG
DNAFISH_twi4_RV	AGCCCCTCTCGGACTATAGT
DNAFISH_twi5_FW	ACCAGACGCCATTTTGCATT
DNAFISH_twi5_RV	TCCGTGTCCTTTGCCACTAT
DNAFISH_twi6_FW	AATGAGTCGAAGGAGCCCAA
DNAFISH_twi6_RV	ATGAACTTCCACGCACACAC
DNAFISH_-181-1_FW	TGCGAAACGATGCTGGAAAT
DNAFISH_-181-1_RV	AGACAGACAAACGGACGGAT
DNAFISH_-181-2_FW	AAGGGGAACGGGACTTACTG
DNAFISH_-181-2_RV	AACAAATCACACCGAGCGAC
DNAFISH_-181-3_FW	CCCCAGCAGTTAAGAGAGCA
DNAFISH_-181-3_RV	GCACCTGTACGACATTTCCA
DNAFISH_-181-4_FW	GCCGAATCCTTGAGTCAAA
DNAFISH_-181-4_RV	CGCTTCTTCACTTTGGCA
DNAFISH_-181-5_FW	TCCATCTCACTTCCACACCC
DNAFISH_-181-5_RV	ATTTCCGCTCCGTTAACCG
DNAFISH_-1.6-1_FW	TTTCGCTTTTCTGTGGGCAA
DNAFISH_-1.6-1_RV	TGAGGCGTTCTAAGTGAGCA
DNAFISH_-1.6-2_FW	CATGTTGCCAGCGATCATCA
DNAFISH_-1.6-2_RV	GCCCTGTTAACACTTTGCCA
DNAFISH_-1.6-3_FW	GCAGCAGTTCTTTCCACAA
DNAFISH_-1.6-3_RV	CTTGTTTGTGAGCAGGCTGT
DNAFISH_-1.6-4_FW	TGCGGACAAAGATCAAGTGC
DNAFISH_-1.6-4_RV	TCTGTCGTGTATGCCGTGTA
DNAFISH_-1.6-5_FW	GCACTAATCCGGAGTCTCGA
DNAFISH_-1.6-5_RV	GCAAGAATGTGGAGCTCTCG

DNAFISH_-1.6-6_FW	CTGGCCCTTTATTCCGAAGC
DNAFISH_-1.6-6_RV	GCATCTCGTTCCTTCTGCAG
DNAFISH_+51-1_FW	TCTAGTTTCCCGCACACAGT
DNAFISH_+51-1_RV	AAGTTGCCAGGCCTATGTCA
DNAFISH_+51-2_FW	CGAATTGATCTGGGTCCTCC
DNAFISH_+51-2_RV	GGTTTACAGACTGTCACGCA
DNAFISH_+51-3_FW	AGGCGTCGTGTTTCATTTCC
DNAFISH_+51-3_RV	TCTATATACCAACGTGCTGAGC
DNAFISH_+51-4_FW	CTTCAGCAAGCAGGACGAAG
DNAFISH_+51-4_RV	TTTGTTCCCCTCGCCGTATA
DNAFISH_+51-5_FW	CTTCCTGAAATGTGGCGCAA
DNAFISH_+51-5_RV	AACTCGTCGTACCATCTGCA
DNAFISH_+51-6_FW	GATGGTGTGCGTCTTGTTGT
DNAFISH_+51-6_RV	CTTGCTGCCAACTCGACTG
DNAFISH_chr3L-1_FW	TTCGGTCTCCCAGATCCAAC
DNAFISH_chr3L-1_RV	CTTCTATGTGTGACCCGTGC
DNAFISH_chr3L-2_FW	GTCGCCGCATTTTCATTGTTG
DNAFISH_chr3L-2_RV	TCTTCTTCTGGGCCGTCAA
DNAFISH_chr3L-3_FW	TCCACTGGCTTTGTTTTCGT
DNAFISH_chr3L-3_RV	CAACTCGCCACAAACCTGAA
DNAFISH_chr3L-4_FW	TGGGCTACACAGGGAGAAAG
DNAFISH_chr3L-4_RV	CTGCTGCTGTTTTGGCCTTA
DNAFISH_chr3L-5_FW	TCGTTGCGAGAACAGAGTCT
DNAFISH_chr3L-5_RV	CGATAAGTGCAGCCTGGAAC
DNAFISH_chr3L-6_FW	GACTGCATCACTTTGAGCCA
DNAFISH_chr3L-6_RV	GAGGGGAAGATTCTACGGCA

**Table 10: List of primers used during this study**

## 1.4 Fly stocks

### *MiMIC fly lines used in this study*

All MiMIC (*Minos* mediated integration cassette) fly lines were ordered from the Bloomington *Drosophila* stock center (BDSC). Each of them contains a gene-trap cassette and the *yellow*<sup>+</sup> marker flanked by two inverted  $\Phi$ C31 integrase *attP* sites at a given location. The sequence of the MiMIC transposable construct can be found under the accession number: GenBank GU370067.1.

BDSC stock number	Genotype	Insertion site
55415	<i>y[1] w[*]; Mi{y[+mDint2]=MIC}MI01218</i>	chr2R:22865023..22865023
38170	<i>y[1] w[*]; Mi{y[+mDint2]=MIC}MI04814</i>	chr2R:21381841..21381841
32829	<i>y[1] w[*]; Mi{y[+mDint2]=MIC}MI02100</i>	chr2R:23097536..23097536
55595	<i>y[1] w[*]; Mi{y[+mDint2]=MIC}MI11229</i>	chr2R:23084945..23084945
55560	<i>y[1] w[*]; Mi{y[+mDint2]=MIC}MI10934</i>	chr3L:6820484..6820484.

**Table 11: List of MiMIC fly lines using in this study**

### *Other fly lines used in this study*

BDSC stock number	Genotype
51324	<i>w[1118]; PBac{y[+mDint2]=vas-Cas9}VK00027</i>
Common lab line	<i>y[*], w[*]</i>
Single balancer lab line	<i>y[*], w[*]; Sco/CyO</i>
Double balancer lab line	<i>w[*]; lf/CyO; Sb/TM3, Ser</i>
766	<i>y[1] w[67c23] P{y[+mDint2]=Crey}1b; sna[Sco]/CyO</i>

**Table 12: List of standard fly lines using in this study**

Fly lines generated during this study

YGH stock number	Genotype	Description
118	<i>w[1118]; twi{attP-3xP3::dsRed-attP}<sup>ΔE3</sup>/CyO</i>	Fly line where the endogenous E3 enhancer has been deleted (called “delta E3”)
126	<i>w[1118]; twi{E3}attP</i>	Fly line where the endogenous E3 enhancer has been deleted and replaced by the E3 sequence in the sens orientation (called “SENS”)
136	<i>w[1118]; twi{In(E3)}attP</i>	Fly line where the endogenous E3 enhancer has been deleted and replaced by the E3 sequence in the anti-sens orientation (called “Anti-Sens”)
54	<i>y[1] w[*]; Mi{E3}MIO1218</i>	Fly line containing the E3 sequence inserted -181 kb away from the promoter of <i>twist</i> (called “-181kb”)
55	<i>y[1] w[*]; Mi{E3}MIO4814</i>	Fly line containing the E3 sequence inserted -1.6 Mb away from the promoter of <i>twist</i> (called “-1.6Mb”)
60	<i>y[1] w[*]; Mi{E3}MIO2100</i>	Fly line containing the E3 sequence inserted +39 kb away from the promoter of <i>twist</i> (called “+39kb”)
61	<i>y[1] w[*]; Mi{E3}MI11229</i>	Fly line containing the E3 sequence inserted +51 kb away from the promoter of <i>twist</i> (called “+51kb”)
62	<i>y[1] w[*]; Mi{E3}MI10934</i>	Fly line containing the E3 sequence inserted on chromosome 3L (called “chr3L”)
51	<i>w[*]; 59C2{attP-3XP3::dsRed}; PBac{y[+mDint2]=vas-Cas9}VK00027</i>	Fly line containing an <i>attP</i> site inserted +7.5 kb away from the promoter of <i>twist</i>
134	<i>w[*]; 59C2{E3}attP; PBac{y[+mDint2]=vas-Cas9}VK00027</i>	Fly line containing the E3 sequence inserted +7.5 kb away from the promoter of <i>twist</i> (called “+7.5kb”)
195	<i>w[*]; twi{attP-3xP3::dsRed-attP}<sup>ΔE3</sup> 59C2{E3}attP; PBac{y[+mDint2]=vas-Cas9}VK00027</i>	Fly line containing the E3 sequence inserted +7.5 kb away from the promoter of <i>twist</i> and where the endogenous E3 has been deleted (called “+7.5kb_DeIE3”)
155	<i>y[1] w[*]; twi{attP-3xP3::dsRed-attP}<sup>ΔE3</sup>; Mi{E3}MI11229</i>	Fly line containing the E3 sequence inserted +51 kb away from the promoter of <i>twist</i> and where the endogenous E3 has been deleted (called “+51kb_DeIE3”)

142	<i>y[1] w[*]; twi{attP-3xP3::dsRed-attP}<sup>ΔE3</sup>/CyO; Mi{E3}MI10934</i>	Fly line containing the E3 sequence inserted at chr3L and where the endogenous E3 has been deleted
-----	---	--

**Table 13: List of fly lines generated during this study**



## 2. Methods

### 2.1 Generation of transgenic fly lines

To generate transgenic fly lines carrying the ectopic insertion of enhancer E3 at different locations and distances from the promoter of *twist*, two methods have been used:

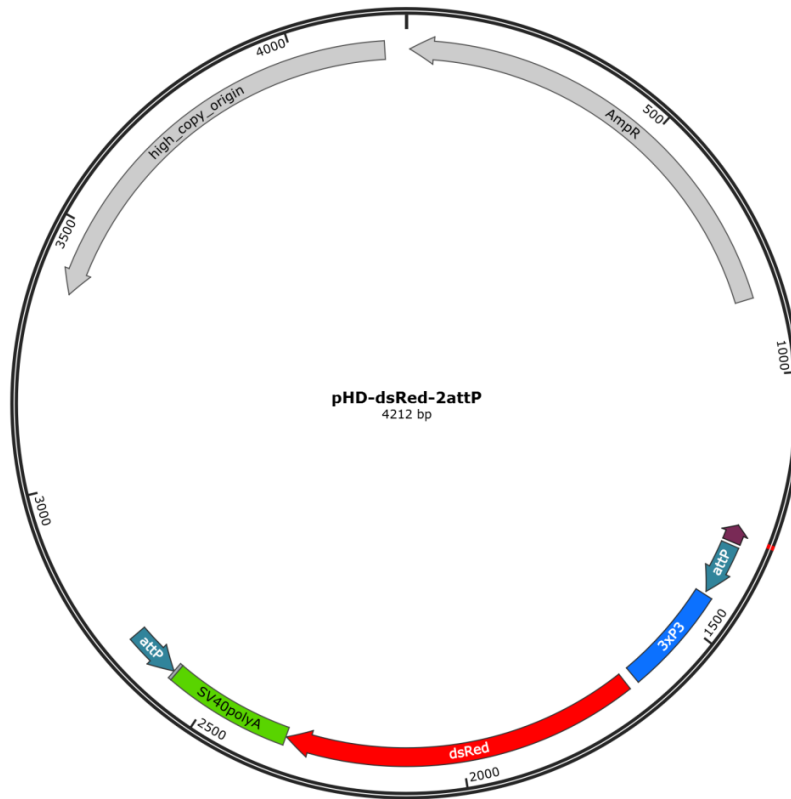
1. Homology-directed repair (HDR) using CRISPR-Cas9 genome editing was used to generate mutant fly lines, each containing a  $\Phi$ C31 integrase *attP* docking site at the desired location. This landing site can then be used to integrate any construct of interest using the  $\Phi$ C31 integrase.
2. Fly lines already containing a *Minos* mediated integration cassette (MiMIC) were obtained from Bloomington. This cassette consists of a transposable construct that enables the insertion of any construct of interest using the  $\Phi$ C31 integrase in a rapid and efficient one-step procedure.

### Plasmids

All plasmids were created using the standard Addgene cloning method with New England Biolabs (NEB) restriction enzymes and T4 DNA ligase (New England Biolabs) or with the NEBuilder HiFi DNA Assembly kit, followed by transformation in *E. coli* and plasmid isolation using the Qiagen midi-prep kit.

#### For the deletion of the endogenous enhancer E3:

The sequence of the  $\Phi$ C31 integrase *attP* landing sites was amplified from the *pJET1.2-STOP-dsRed* vector, using either the SV788 and SV789 or the SV790 and SV792 primers (Table 8) which add on each side of the PCR product either a BsiWI restriction site or an AgeI and a SpeI restriction site respectively. These PCR products were then sequentially cloned into the *pHD-dsRed* vector by digesting it with the same restriction enzymes. The obtained vector was called *pHD-dsRed-2attP* (Figure 31). It contains a *dsRed* fluorescent cassette driven by the *3xP3* promoter (red fluorescence in the eye), flanked on each side by a  $\Phi$ C31 integrase *attP* site. It also contains a *LoxP* site just downstream of the first *attP* site.



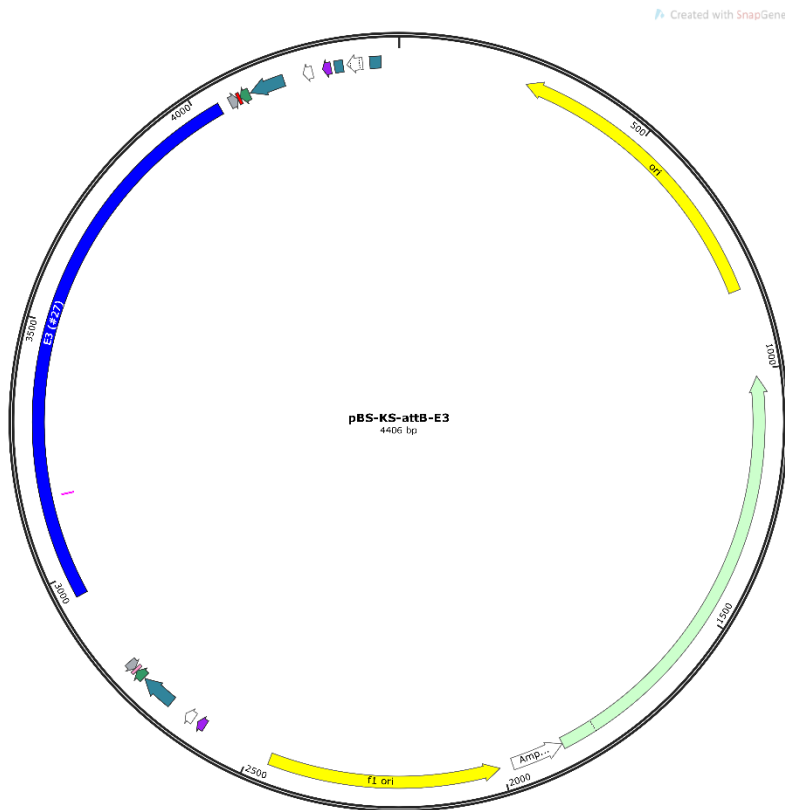
**Figure 31: Map of the *pHD-dsRed-2attP* vector generated in this study**

The *pHD-dsRed-2attP* was then used to clone the two homology arms that correspond to a ~ 1 kb region surrounding the E3 enhancer. Both homology arms were amplified from genomic DNA extracted from fly w[1118]; PBac{y[+mDint2]=vas-Cas9}VK00027. Homology arm 1 was amplified using the SV797 and SV798 primers which add a NheI restriction site on each side of the PCR product. Homology arm 2 was amplified using the SV799 and SV800 primers which add a SpeI restriction site on one side and a PstI restriction site on the other side of the PCR product. These PCR products were then sequentially cloned into the *pHD-dsRed-2attP* vector by digesting it with the same restriction enzymes. The obtained vector was called *pHD-dsRed-2attP-HA1&2*.

The gRNAs were designed using the flyCRISPR target finder (Gratz et al., 2014) and cloned in the *pU6-BbsI-gRNA* vector after digestion with the BbsI enzyme following the recommendations from the flyCRISPR website. The obtained vectors were called *pU6-BbsI-gRNA5'* and *pU6-BbsI-gRNA3'*.

For the Inverted Orientation of the endogenous enhancer E3:

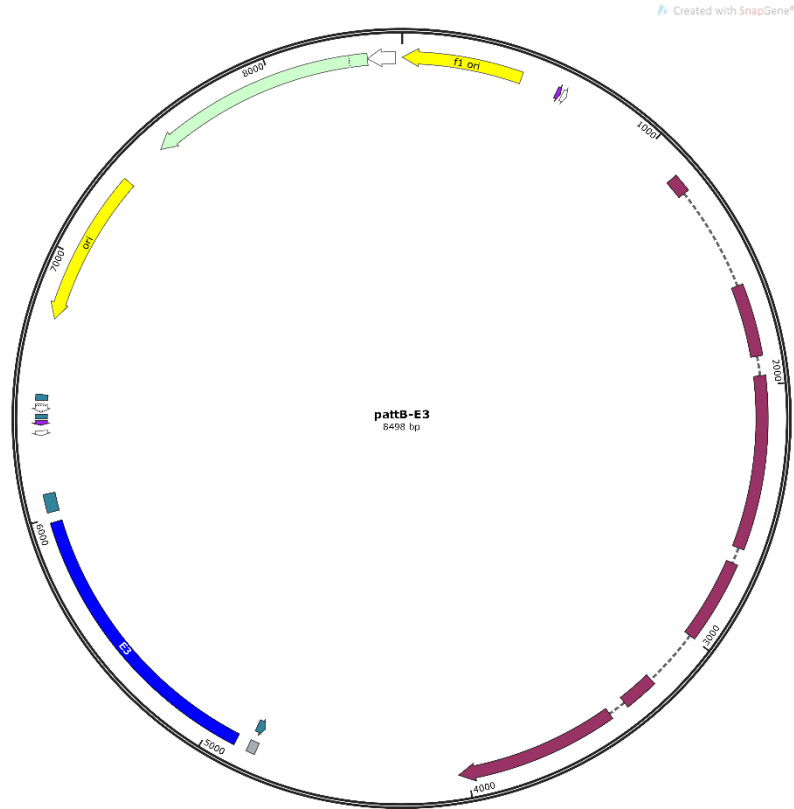
The E3 sequence was amplified from genomic DNA extracted from the fly line w[1118]; PBac[y[+mDint2]=vas-Cas9]VK00027 using the PCR primers SV908 and SV858, and integrated into the *pBS-KS-attB1-2-PT-SA-SD-0-2xTY1-V5* vector digested with the XbaI and HindIII enzymes using the NEB HiFi DNA Assembly kit. The obtained vector was called *pBS-KS-attB-E3* (Figure 32).



**Figure 32: Map of the *pBS-KS-attB-E3* vector generated in this study**

For the ectopic integration of enhancer E3 at + 7.5 kb away from the *twist* promoter:

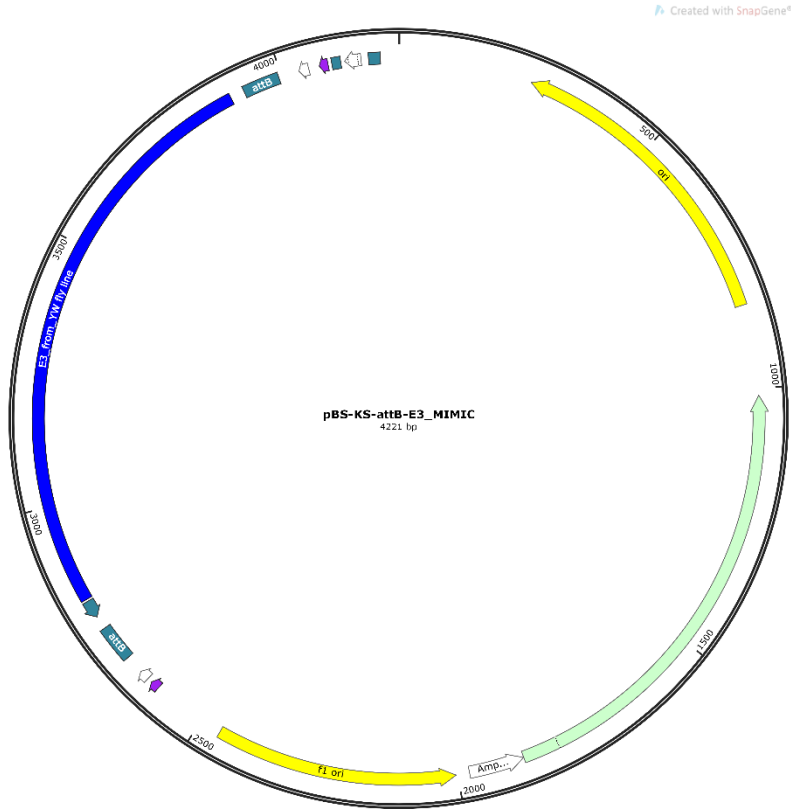
The fly line was generated in two steps: first, CRISPR-Cas9 genome editing was used to generate a fly line containing a  $\Phi$ C31 integrase *attP* landing site + 7.5 kb away from the *twist* promoter. For this purpose, the *pHD-dsRed-attP* vector was used to clone the two homology arms that correspond to a ~ 1 kb region surrounding the insertion site. Both homology arms were amplified from genomic DNA extracted from fly w[1118]; PBac{y[+mDint2]=vas-Cas9}VK00027. Homology arm 1 was amplified using the HA1\_FW and HA1\_RV primers which add an AarI restriction site on each side of the PCR product. Homology arm 2 was amplified using the HA2\_FW and HA2\_RV primers which add a SapI restriction site on each side of the PCR product. These PCR products were then sequentially cloned into the *pHD-dsRed-attP* vector by digesting it with the same restriction enzymes. The obtained vector was called *pHD-dsRed-attP-HA1&2*. The gRNAs were designed using the flyCRISPR target finder (Gratz et al., 2014) and cloned in the *pU6-BbsI-gRNA* vector after digestion with the BbsI enzyme following the recommendations from the flyCRISPR website. The obtained vector was called *pU6-BbsI-gRNA*. Second, the resulting fly line was used to insert the E3 sequence. The E3 sequence was amplified from genomic DNA extracted from the fly line *y[\*], w[\*]* using the PCR primers FW\_KpnI and RV\_XhoI which add a KpnI restriction site on one side and an XhoI restriction site on the other side. This PCR product was then cloned into the *pattB* vector digested with the same restriction enzymes (Note: we first deleted the *LoxP* sequence originally present in this plasmid). The obtained vector was called *pattB-E3* (Figure 33).



**Figure 33: Map of the *pattB-E3* vector generated in this study**

For the ectopic integration of enhancer E3 at -1.6 Mb, -181 kb, +39 kb, +51 kb away from the *twist* promoter and on chromosome 3L:

The E3 sequence was amplified from genomic DNA extracted from the  $\gamma[*]$ ,  $w[*]$  fly line using the PCR primers loxp\_28FW\_E3\_Alexia and E3\_HindIII\_R which add an XbaI restriction site on one side and a HindIII restriction site on the other side. This PCR product was then cloned into the *pBS-KS-attB1-2-PT-SA-SD-0-2xTY1-V5* vector digested with the same restriction enzymes. The final vector was called *pBS-KS-attB-E3\_MIMIC* (Figure 34).



**Figure 34: Map of the *pBS-KS-attB-E3\_MIMIC* vector generated in this study**

### *Injections and fly husbandry*

Non-dechorionated pre-blastoderm embryos of the appropriate fly line were injected with a FemtoJet 4i electronic microinjector (Eppendorf) using self-pulled glass needles (Model P-1000 Micropipette Puller - Sutter Instrument). All injected plasmids were purified using the Qiagen Plasmid Midi kit and eluted in H<sub>2</sub>O. The plasmids were diluted in injection buffer at the following concentrations:

- For the generation of the fly line deleted for the endogenous E3: a mix containing 250ng/ul of *pU6-gRNA* plasmid and 500ng/ul of donor plasmid was injected in the w[1118]; PBac{y[+mDint2]=vas-Cas9}VK00027 line.

- For the fly lines carrying E3 in either orientation: a mix containing 500ng/ul of donor plasmid and 100ng/ul of helper plasmid encoding the  $\Phi$ C31 integrase (*p3xP3-EGFP.vas-int.NLS*) was injected in the fly line where the endogenous E3 sequence was deleted.
- For the generation of the ectopic E3 insertions at + 7.5 kb away from the *twist* promoter: a mix containing 250ng/ul of *pU6-gRNA* plasmids and 500ng/ul of donor plasmid was first injected in the *w[1118]; PBac{y[+mDint2]=vas-Cas9}VK00027* line. The resulting fly line was then used to inject a mix containing 500ng/ul of donor plasmid and 100ng/ul of helper plasmid encoding the  $\Phi$ C31 integrase (*p3xP3-EGFP.vas-int.NLS*). The resulting fly line was then crossed to the *y[1] w[67c23] P{y[+mDint2]=Crey}1b; sna[Sco]/CyO* fly line to delete the *dsRed* cassette.
- For the generation of the ectopic E3 insertions at -1.6 Mb, -181 kb, +39 kb, +51 kb away from the *twist* promoter and on chromosome 3L: a mix containing 500ng/ul of donor plasmid and 100ng/ul of helper plasmid encoding the  $\Phi$ C31 integrase (*p3xP3-EGFP.vas-int.NLS*) was injected in the appropriate MIMIC fly line.
- For the deletion of the endogenous E3 sequence from the lines where E3 was ectopically inserted -1.6 Mb, -181 kb, +39 kb, +51 kb away from the *twist* promoter: a mix containing 250ng/ul of *pU6-gRNA* plasmids and 500ng/ul of donor plasmid was injected in each of the fly lines which were previously crossed with the *w[1118]; PBac{y[+mDint2]=vas-Cas9}VK00027* line. The deletion of the endogenous E3 sequence in the line where E3 was ectopically inserted on chromosome 3L was simply obtained by fly crosses.
- For the deletion of the endogenous E3 sequence from the lines where E3 was ectopically inserted + 7.5 kb away from the *twist* promoter: a mix containing 250ng/ul of *pU6-gRNA* plasmids and 500ng/ul of donor plasmid was injected in each of the fly lines which were previously crossed with the *w[1118]; PBac{y[+mDint2]=vas-Cas9}VK00027* line.

Flies were raised on a standard agar medium and grown at 25°C under standard conditions.

## 2.2 4C-sequencing

### 4C library generation

4C templates were generated as previously described with minor modifications (Krijger et al., 2020b). Freshly hatched adults of the appropriate genotype were placed in embryo collection vials with standard apple cap plates. *Drosophila* embryos were collected (after 3 pre-lays of 1 hour) at 2–5 h after egg-lay and covalently crosslinked in 1.8% formaldehyde for 15 min at room temperature. Nuclear extraction was carried as described previously (Ghavi-Helm et al., 2014a). About 30 million nuclei were used for each 4C template preparation using MboI and NlaIII as the first and second restriction enzymes, respectively.

4C templates were amplified from 320 ng of 4C template using the primers listed in Table 11. Each primer contains a 4 to 8 nucleotides “shift” sequence (red), which is used to allow optimal base-pair diversity at the beginning of the read after multiplexing. The viewpoint is located at coordinates 2R:23,044,042..23,044,236 upstream for the *twist* promoter.

4C primer	Sequence	NEB index	Used for library
<b>Tw1_FW</b>	<b>HBVV</b> TACGTGCACCAAAAAGTTTCTT	ATCACG	-181kb-rep1 and +51kb_DeIE3-rep1
<b>Tw1_RV</b>	<b>Bddb</b> AAAATGGTCGTCAAAGCGC		
<b>Tw1_FW</b>	<b>HBVV</b> TACGTGCACCAAAAAGTTTCTT	CGATGT	-181kb-rep2 and +51kb_DeIE3-rep2
<b>Tw1_RV</b>	<b>Bddb</b> AAAATGGTCGTCAAAGCGC		
<b>Tw2_FW</b>	<b>BHBVV</b> TACGTGCACCAAAAAGTTTCTT	TTAGGC	-1.6Mb-rep1 and +7.5kb-rep1
<b>Tw2_RV</b>	<b>VBddb</b> AAAATGGTCGTCAAAGCGC		
<b>Tw2_FW</b>	<b>BHBVV</b> TACGTGCACCAAAAAGTTTCTT	TGACCA	-1.6Mb-rep2 and +7.5kb-rep2
<b>Tw2_RV</b>	<b>VBddb</b> AAAATGGTCGTCAAAGCGC		



<b>Tw3_FW</b>	<b>HBHBVV</b> TACGTGCACCAAAAAGTTTCTT	ACAGTG	+51kb-rep1 and +7.5kb_DelE3-rep1
<b>Tw3_RV</b>	<b>HVBDDDB</b> AAAATGGTCGTCAAAGCGC		
<b>Tw3_FW</b>	<b>HBHBVV</b> TACGTGCACCAAAAAGTTTCTT	GCCAAT	+51kb-rep2 and +7.5kb_DelE3-rep2
<b>Tw3_RV</b>	<b>HVBDDDB</b> AAAATGGTCGTCAAAGCGC		
<b>Tw4_FW</b>	<b>DHBHBVV</b> TACGTGCACCAAAAAGTTTCTT	CAGATC	+39kb-rep1 and AS-rep1
<b>Tw4_RV</b>	<b>VHVBDDDB</b> AAAATGGTCGTCAAAGCGC		
<b>Tw4_FW</b>	<b>DHBHBVV</b> TACGTGCACCAAAAAGTTTCTT	ACTTGA	+39kb-rep2 and AS-rep2
<b>Tw4_RV</b>	<b>VHVBDDDB</b> AAAATGGTCGTCAAAGCGC		
<b>Tw5_FW</b>	<b>HDHBHBVV</b> TACGTGCACCAAAAAGTTTCTT	GATCAG	chr3L-rep1 and SENS-rep1
<b>Tw5_RV</b>	<b>VVHVBDDDB</b> AAAATGGTCGTCAAAGCGC		
<b>Tw5_FW</b>	<b>HDHBHBVV</b> TACGTGCACCAAAAAGTTTCTT	TAGCTT	chr3L-rep2 and SENS-rep2
<b>Tw5_RV</b>	<b>VVHVBDDDB</b> AAAATGGTCGTCAAAGCGC		

**Table 14: List of 4C primers**

The PCR product was purified using SPRIselect beads and 80 ng of each PCR product was used to generate the final libraries using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB). The libraries were indexed for multiplexing using NEBNext multiplex oligos kit for Illumina (Index primers set 1) (See Table 12). A total of 20 libraries were generated, with two independent biological replicates for each sample. The libraries were multiplexed in 2 sets of 10 and sequenced on a NovaSeq S4 sequencer (Illumina) by the Novogene company using 150-bp paired-end reads. To ensure sufficient coverage, each sample has at least 10 million reads.

#### 4C data analysis

The quality of the 4C-seq data was confirmed using the FastQC software (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Adapter sequences were trimmed using TrimGalore ([https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)), and the 5' shift sequence and the primer sequence was trimmed up to the location of the first restriction site (the restriction enzyme cutting site was kept) using Cutadapt version 2.10 (Martin, 2011). As sequencing was performed in paired-end mode, the fastq files corresponding to each pair were merged into a single fastq file. Finally, as the read length is relatively long, it is possible to observe reads that result from multiple ligation events between different regions of the genomes. As such reads will not map efficiently to the genome, we scanned the reads from the 5' side for the presence of the first, then of the second restriction site, and trimmed the sequence after this location. As a consequence, the reads typically start with the sequence of the first restriction enzyme and have either the sequence of the first or the second restriction site at their 3' end. The trimmed reads were then aligned to a custom genome, generated using the reform Python tool (<https://github.com/gencorefacility/reform>), which consists of the *D. melanogaster* reference genome (dm6) where the endogenous E3 enhancer sequence has been deleted and re-introduced at the appropriate location. Six different custom genomes were thus created for each insertion site. Alignment was performed using Bowtie2 version 2.3.5.1 (Langmead and Salzberg, 2012). We generated a "4C valid fragment" library based on the recognition sequences of the first restriction enzyme within each custom genome, and only kept the subset of mapped reads that overlapped these valid fragments. The libraries were then normalized by scaling using the read per million method and transformed into coverage bedgraph files. 4C data were plotted using pyGenomeTracks version 3.6 (Ramírez et al., 2018) (Lopez-Delisle et al., 2021).

### 2.3 3D DNA-FISH

3D DNA FISH was performed as previously described (Szabo et al., 2021), on dechorionated embryos from a mix of overday and overnight collections fixed for 25 min with 4% formaldehyde. DNA FISH probes were constructed by PCR using the probes listed in Table 8 as previously described (Szabo et al., 2021). FISH probes were labeled using the FISH Tag DNA Multicolor kit (Alexa Fluor 488 dye for the *twist* promoter and Alexa Fluor 555 dye for the ectopic insertion sites) (Life Technologies). The slides were mounted using ProLong Gold antifade reagent with DAPI (Life Technologies) and imaged on a Leica SP8 confocal microscope. 3D stacks were collected, and relative distances between FISH signals were analyzed in at least 500 nuclei from 3 to 4 independent embryos using the Imaris software (Bitplane).

### 2.4 RNA-sequencing

#### *RNA-seq library generation*

Freshly hatched adults of the appropriate genotype were placed in embryo collection vials with standard apple cap plates. *Drosophila* embryos were collected (after 3 pre-lays of 1 hour) at 2–5 h after egg-lay and snap-frozen in RA1 lysis buffer with  $\beta$ -mercaptoethanol (NUCLEOSPIN RNA kit - Macherey-Nagel). The embryos were then homogenized in the lysis buffer with a Cordless Motor for Pellet Mix and pestles (VWR) on ice, and the RNA was extracted following the NUCLEOSPIN RNA kit (Macherey-Nagel) protocol. RNA concentration was measured using the Qubit™ RNA HS Assay kit and RNA quality was analyzed using the RNA TapeStation kit (Agilent). Strand-specific RNA-sequencing libraries were generated by the IGFL Sequencing platform (PSI) using the SENSE mRNA-Seq Library Prep kit V2 (Lexogen) with the i7 6nt Index set (7001-7096) according to the manufacturers' instructions. The samples were multiplexed and sequenced on a NextSeq500 sequencer (Illumina) using 75-bp single-end reads. Three independent biological replicates were performed for each sample.

### *RNA-seq data analysis*

The quality of the RNA-seq data was confirmed using the FastQC software (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Adapter sequences were trimmed using TrimGalore ([https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)), before mapping to the *D. melanogaster* reference genome (dm6) using STAR version 2.5.4b (Dobin et al., 2013). Quantification of reads corresponding to each gene was performed by featureCounts version 2.0.1 (Liao et al., 2014). Normalization and differential gene expression analysis was performed using DESeq2 version 1.32.0 (Love et al., 2014).



# References:

- Abdulghani, M., Jain, A., and Tuteja, G. (2019). Genome-wide identification of enhancer elements in the placenta. *Placenta* 79, 72–77.
- Akbari, O.S., Bae, E., Johnsen, H., Villaluz, A., Wong, D., and Drewell, R.A. (2008). A novel promoter-tethering element regulates enhancer-driven gene expression at the bithorax complex in the *Drosophila* embryo. *Development* 135, 123–131.
- Alexander, R.P., Fang, G., Rozowsky, J., Snyder, M., and Gerstein, M.B. (2010). Annotating non-coding regions of the genome. *Nat Rev Genet* 11, 559–571.
- Andrey, G., and Mundlos, S. (2017). The three-dimensional genome: regulating gene expression during pluripotency and development. *Development* 144, 3646–3658.
- van Arensbergen, J., van Steensel, B., and Bussemaker, H.J. (2014). In search of the determinants of enhancer–promoter interaction specificity. *Trends in Cell Biology* 24, 695–702.
- Atchison, M.L. (1988). Enhancers: Mechanisms of Action and Cell Specificity. *Annual Review of Cell Biology* 4, 127–153.
- Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a  $\beta$ -globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299–308.
- Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* 33, 729–740.
- Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in *Drosophila*. *Cell* 144, 214–226.
- Baylies, M.K., and Bate, M. (1996). twist: A Myogenic Switch in *Drosophila*. *Science* 272, 1481–1484.
- Benabdallah, N.S., Williamson, I., Illingworth, R.S., Kane, L., Boyle, S., Sengupta, D., Grimes, G.R., Therizols, P., and Bickmore, W.A. (2019). Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Molecular Cell* 76, 473-484.e7.
- Blackwood, E.M., and Kadonaga, J.T. (1998). Going the Distance: A Current View of Enhancer Action. *Science* 281, 60–63.
- Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E.M. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* 44, 148–156.
- Brown, W., and Clancy, S. (2008). Translation: DNA to mRNA to Protein. Undefined.

- Brown, J.M., Roberts, N.A., Graham, B., Waithe, D., Lagerholm, C., Telenius, J.M., De Ornellas, S., Oudelaar, A.M., Scott, C., Szczerbal, I., et al. (2018). A tissue-specific self-interacting chromatin domain forms independently of enhancer-promoter interactions. *Nat Commun* *9*, 3849.
- Bulger, M., and Groudine, M. (2011). Functional and Mechanistic Diversity of Distal Transcription Enhancers. *Cell* *144*, 327–339.
- Butler, J.E.F. (2001). Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes & Development* *15*, 2515–2519.
- Caglio, G., Triglia, E.T., and Pombo, A. (2017). Keep Them Close: PRC2 Poises Enhancer-Promoter Interactions at Anterior Neuronal Genes. *Cell Stem Cell* *20*, 573–575.
- Calhoun, V.C., and Levine, M. (2003). Long-range enhancer–promoter interactions in the Scr-Antp interval of the *Drosophila* Antennapedia complex. *PNAS* *100*, 9878–9883.
- Calhoun, V.C., Stathopoulos, A., and Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila* Antennapedia complex. *Proc Natl Acad Sci U S A* *99*, 9243–9247.
- Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how and why? *Mol Cell* *49*.
- Cavalheiro, G.R., Pollex, T., and Furlong, E.E. (2021). To loop or not to loop: what is the role of TADs in enhancer function and gene regulation? *Current Opinion in Genetics & Development* *67*, 119–129.
- Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma’ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* *14*, 128.
- Cheutin, T., and Cavalli, G. (2019). The multiscale effects of polycomb mechanisms on 3D chromatin folding. *Critical Reviews in Biochemistry and Molecular Biology* *54*, 399–417.
- Cho, W.-K., Jayanth, N., English, B.P., Inoue, T., Andrews, J.O., Conway, W., Grimm, J.B., Spille, J.-H., Lavis, L.D., Lionnet, T., et al. (2016). RNA Polymerase II cluster dynamics predict mRNA output in living cells. *Elife* *5*.
- Cooper, G.M. (2000). *Eukaryotic RNA Polymerases and General Transcription Factors. The Cell: A Molecular Approach. 2nd Edition.*
- Core, L.J., Martins, A.L., Danko, C.G., Waters, C., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* *46*, 1311–1320.
- Cubeñas-Potts, C., Rowley, M.J., Lyu, X., Li, G., Lei, E.P., and Corces, V.G. (2017). Different enhancer classes in *Drosophila* bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Research* *45*, 1714–1730.

- Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing Chromosome Conformation. *Science* 295, 1306–1311.
- Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149, 1233–1244.
- Despang, A., Schöpflin, R., Franke, M., Ali, S., Jerković, I., Paliou, C., Chan, W.-L., Timmermann, B., Wittler, L., Vingron, M., et al. (2019). Functional dissection of the Sox9 – Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat Genet* 51, 1263–1271.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Dong, Z., Wang, H., Chen, H., Jiang, H., Yuan, J., Yang, Z., Wang, W.-J., Xu, F., Guo, X., Cao, Y., et al. (2018). Identification of balanced chromosomal rearrangements previously unknown among participants in the 1000 Genomes Project: implications for interpretation of structural variation in genomes and the future of clinical cytogenetics. *Genetics in Medicine* 20, 697.
- Espinola, S.M., Götz, M., Bellec, M., Messina, O., Fiche, J.-B., Houbron, C., Dejean, M., Reim, I., Cardozo Gizzi, A.M., Lagha, M., et al. (2021). Cis -regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early *Drosophila* development. *Nat Genet* 53, 477–486.
- Farley, E., and Levine, M. (2012). HOT DNAs: a novel class of developmental enhancers. *Genes Dev.* 26, 873–876.
- Field, A., and Adelman, K. (2020). Evaluating Enhancer Function and Transcription. *Annu. Rev. Biochem.* 89, 213–234.
- Flöttmann, R., Wagner, J., Kobus, K., Curry, C.J., Savarirayan, R., Nishimura, G., Yasui, N., Spranger, J., Esch, H.V., Lyons, M.J., et al. (2015). Microdeletions on 6p22.3 are associated with mesomelic dysplasia Savarirayan type. *Journal of Medical Genetics* 52, 476–483.
- Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., Chan, W.-L., et al. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* 538, 265–269.
- Fudenberg, G., and Imakaev, M. (2017). FISH-ing for captured contacts: towards reconciling FISH and 3C. *Nat Methods* 14, 673–678.
- Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. *GENES IN DEVELOPMENT* 6.
- Furlong, E.E.M., Andersen, E.C., Null, B., White, K.P., and Scott, M.P. (2001). Patterns of Gene Expression During *Drosophila* Mesoderm Development. *Science* 293, 1629–1633.



Galupa, R., Nora, E.P., Worsley-Hunt, R., Picard, C., Gard, C., van Bemmelen, J.G., Servant, N., Zhan, Y., El Marjou, F., Johanneau, C., et al. (2020). A Conserved Noncoding Locus Regulates Random Monoallelic Xist Expression across a Topological Boundary. *Molecular Cell* 77, 352–367.e8.

Gambetta, M.C., and Furlong, E.E.M. (2018). The Insulator Protein CTCF Is Required for Correct Hox Gene Expression, but Not for Embryonic Development in *Drosophila*. *Genetics* 210, 129–136.

Ghavi-Helm, Y. (2019). Functional consequences of chromosomal rearrangements on gene expression: not so deleterious after all? *Journal of Molecular Biology*.

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E.M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* 512, 96–100.

Ghavi-Helm, Y., Jankowski, A., Meiers, S., Viales, R.R., Korb, J.O., and Furlong, E.E.M. (2019). Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat Genet* 51, 1272–1282.

Giorgetti, L., and Heard, E. (2016). Closing the loop: 3C versus DNA FISH. *Genome Biology* 17, 215.

Giorgio, E., Robyr, D., Spielmann, M., Ferrero, E., Di Gregorio, E., Imperiale, D., Vaula, G., Stamoulis, G., Santoni, F., Atzori, C., et al. (2015). A large genomic deletion leads to enhancer adoption by the lamin B1 gene: a second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). *Hum Mol Genet* 24, 3143–3154.

Golov, A.K., Abashkin, D.A., Kondratyev, N.V., Razin, S.V., Gavrillov, A.A., and Golimbet, V.E. (2020). A modified protocol of Capture-C allows affordable and flexible high-resolution promoter interactome analysis. *Sci Rep* 10, 15491.

Gratz, S.J., Ukken, F.P., Rubinstein, C.D., Thiede, G., Donohue, L.K., Cummings, A.M., and O'Connor-Giles, K.M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics* 196, 961–971.

Gu, B., Swigut, T., Spencley, A., Bauer, M.R., Chung, M., Meyer, T., and Wysocka, J. (2018). Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* 359, 1050–1055.

Gurudatta, B.V., and Corces, V.G. (2009). Chromatin insulators: lessons from the fly. *Brief Funct Genomic Proteomic* 8, 276–282.

Hamm, D.C., and Harrison, M.M. (2018). Regulatory principles governing the maternal-to-zygotic transition: insights from *Drosophila melanogaster*. *Open Biol* 8, 180183.

Hanssen, L.L.P., Kassouf, M.T., Oudelaar, A.M., Biggs, D., Preece, C., Downes, D.J., Gosden, M., Sharpe, J.A., Sloane-Stanley, J.A., Hughes, J.R., et al. (2017). Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nat Cell Biol* 19, 952–961.

Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev* 32, 26–41.

Hrdlickova, B., de Almeida, R.C., Borek, Z., and Withoff, S. (2014). Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1842, 1910–1922.

Hsieh, T.-H.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Rando, O.J., Tjian, R., and Darzacq, X. (2020). Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Molecular Cell* 78, 539-553.e8.

Hsieh, T.-H.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Darzacq, X., and Tjian, R. (2021). Enhancer-promoter interactions and transcription are maintained upon acute loss of CTCF, cohesin, WAPL, and YY1. *BioRxiv* 2021.07.14.452365.

Hua, P., Badat, M., Hanssen, L.L.P., Hentges, L.D., Crump, N., Downes, D.J., Jeziorska, D.M., Oudelaar, A.M., Schwessinger, R., Taylor, S., et al. (2021). Defining genome architecture at base-pair resolution. *Nature* 595, 125–129.

Hug, C.B., Grimaldi, A.G., Kruse, K., and Vaquerizas, J.M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* 169, 216-228.e19.

Ing-Simmons, E., Vaid, R., Bing, X.Y., Levine, M., Mannervik, M., and Vaquerizas, J.M. (2021). Independence of chromatin conformation and gene regulation during *Drosophila* dorsoventral patterning. *Nat Genet* 53, 487–499.

Jennings, B.H. (2011). *Drosophila* – a versatile model in biology & medicine. *Materials Today* 14, 190–195.

Jiang, J., Kosman, D., Ip, Y.T., and Levine, M. (1991). The dorsal morphogen gradient regulates the mesoderm determinant twist in early *Drosophila* embryos. *Genes Dev.* 5, 1881–1891.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* 503, 290–294.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and Cohesin Connect Gene Expression and Chromatin Architecture. *Nature* 467, 430–435.

Kim, J., and Dean, A. (2021). Enhancers navigate the three-dimensional genome to direct cell fate decisions. *Current Opinion in Structural Biology* 71, 101–109.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187.

Koenecke, N., Johnston, J., Gaertner, B., Natarajan, M., and Zeitlinger, J. (2016). Genome-wide identification of *Drosophila* dorso-ventral enhancers by differential histone acetylation analysis. *Genome Biology* 17, 196.

Kolovos, P., Knoch, T.A., Grosveld, F.G., Cook, P.R., and Papantonis, A. (2012). Enhancers and silencers: an integrated and simple model for their function. *Epigenetics & Chromatin* 5, 1.

Krietenstein, N., Abraham, S., Venev, S.V., Abdennur, N., Gibcus, J., Hsieh, T.-H.S., Parsi, K.M., Yang, L., Maehr, R., Mirny, L.A., et al. (2020). Ultrastructural Details of Mammalian Chromosome Architecture. *Molecular Cell* 78, 554-565.e7.

Krijger, P.H.L., Geeven, G., Bianchi, V., Hilvering, C.R.E., and de Laat, W. (2020a). 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* 170, 17–32.

Krivega, I., and Dean, A. (2012). Enhancer and promoter interactions—long distance calls. *Current Opinion in Genetics & Development* 22, 79–85.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 44, W90-97.

Kvon, E.Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., and Stark, A. (2014). Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature* 512, 91–95.

Kyrchanova, O., Maksimenko, O., Stakhov, V., Ivlieva, T., Parshikov, A., Studitsky, V.M., and Georgiev, P. (2013). Effective Blocking of the White Enhancer Requires Cooperation between Two Main Mechanisms Suggested for the Insulator Function. *PLOS Genetics* 9, e1003606.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359.

Laugsch, M., Bartusel, M., Rehim, R., Alirzayeva, H., Karaolidou, A., Crispatsu, G., Zentis, P., Nikolic, M., Bleckwehl, T., Kolovos, P., et al. (2019a). Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs. *Cell Stem Cell* 24, 736-752.e12.

Le Dily, F., Baù, D., Pohl, A., Vicent, G.P., Serra, F., Soronellas, D., Castellano, G., Wright, R.H.G., Ballare, C., Filion, G., et al. (2014). Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev* 28, 2151–2162.

Leptin, M. (1991). twist and snail as positive and negative regulators during *Drosophila* mesoderm development. *Genes Dev.* 5, 1568–1576.

Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., Beer, P. de, Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and Graaff, E. de (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* 12, 1725–1735.

- Lettice, L.A., Daniels, S., Sweeney, E., Venkataraman, S., Devenney, P.S., Gautier, P., Morrison, H., Fantes, J., Hill, R.E., and FitzPatrick, D.R. (2011). Enhancer-adoption as a mechanism of human developmental disease. *Hum Mutat* *32*, 1492–1499.
- Levine, M. (2010). Transcriptional Enhancers in Animal Development and Evolution. *Curr Biol* *20*, R754–R763.
- Lewis, M.W., Li, S., and Franco, H.L. (2019). Transcriptional control by enhancers and enhancer RNAs. *Transcription* *10*, 171–186.
- Li, X., and Noll, M. (1994). Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the *Drosophila* embryo. *EMBO J* *13*, 400–406.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* *30*, 923–930.
- Lim, B., and Levine, M.S. (2021). Enhancer-promoter communication: hubs or loops? *Current Opinion in Genetics & Development* *67*, 5–9.
- Lim, B., Heist, T., Levine, M., and Fukaya, T. (2018). Visualization of Transvection in Living *Drosophila* Embryos. *Mol Cell* *70*, 287-296.e6.
- Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J., and Axel, R. (2006). Interchromosomal Interactions and Olfactory Receptor Choice. *Cell* *126*, 403–413.
- Lopez-Delisle, L., Rabbani, L., Wolff, J., Bhardwaj, V., Backofen, R., Grüning, B., Ramírez, F., and Manke, T. (2021). pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* *37*, 422–423.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* *15*, 550.
- Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015a). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* *161*, 1012–1025.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* *17*, 10–12.
- McArthur, E., and Capra, J.A. (2021). Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *The American Journal of Human Genetics* *108*, 269–283.
- Mendenhall, E.M., Williamson, K.E., Reyon, D., Zou, J.Y., Ram, O., Joung, J.K., and Bernstein, B.E. (2013). Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol* *31*, 1133–1136.

- Mercola, M., Wang, X.F., Olsen, J., and Calame, K. (1983). Transcriptional enhancer elements in the mouse immunoglobulin heavy chain locus. *Science* *221*, 663–665.
- Merli, C., Bergstrom, D.E., Cygan, J.A., and Blackman, R.K. (1996). Promoter specificity mediates the independent regulation of neighboring genes. *Genes Dev.* *10*, 1260–1270.
- Moretti, C., Stévant, I., and Ghavi-Helm, Y. (2019). 3D genome organisation in *Drosophila*. *Briefings in Functional Genomics* elz029.
- Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* *502*, 59–64.
- Nègre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C.A., Henikoff, J.G., Feng, X., Ahmad, K., Russell, S., White, R.A.H., et al. (2010). A Comprehensive Map of Insulator Elements for the *Drosophila* Genome. *PLOS Genetics* *6*, e1000814.
- Nègre, N., Brown, C.D., Ma, L., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R., et al. (2011). A cis-regulatory map of the *Drosophila* genome. *Nature* *471*, 527–531.
- Noordermeer, D., and Laats, W. de (2008). Joining the loops:  $\beta$ -Globin gene regulation. *IUBMB Life* *60*, 824–833.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., Berkum, N.L. van, Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* *485*, 381–385.
- Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* *169*, 930-944.e22.
- Northcott, P.A., Lee, C., Zichner, T., Stütz, A.M., Erkek, S., Kawauchi, D., Shih, D.J.H., Hovestadt, V., Zapatka, M., Sturm, D., et al. (2014). Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* *511*, 428–434.
- Ogiyama, Y., Schuettengruber, B., Papadopoulos, G.L., Chang, J.-M., and Cavalli, G. (2018). Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during *Drosophila* Development. *Molecular Cell* *71*, 73-88.e5.
- Oh, S., Shao, J., Mitra, J., Xiong, F., D’Antonio, M., Wang, R., Garcia-Bassets, I., Ma, Q., Zhu, X., Lee, J.-H., et al. (2021). Enhancer release and retargeting activates disease-susceptibility genes. *Nature* 1–6.
- Ohtsuki, S., Levine, M., and Cai, H.N. (1998a). Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. *Genes & Development* *12*, 547–556.
- Ohtsuki, S., Levine, M., and Cai, H.N. (1998b). Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. *Genes Dev* *12*, 547–556.

Oluwadare, O., Highsmith, M., and Cheng, J. (2019). An Overview of Methods for Reconstructing 3-D Chromosome and Genome Structures from Hi-C Data. *Biol Proced Online* 21, 7.

Orphanides, G., and Reinberg, D. (2002). A Unified Theory of Gene Expression. *Cell* 108, 439–451.

Padeken, J., Mendiburo, M.J., Chlamydas, S., Schwarz, H.-J., Kremmer, E., and Heun, P. (2013). The nucleoplasmin homolog NLP mediates centromere clustering and anchoring to the nucleolus. *Mol Cell* 50, 236–249.

Pan, D.J., Huang, J.D., and Courey, A.J. (1991). Functional analysis of the *Drosophila* twist promoter reveals a dorsal-binding ventral activator region. *Genes Dev.* 5, 1892–1901.

Panigrahi, A., and O'Malley, B.W. (2021). Mechanisms of enhancer action: the known and the unknown. *Genome Biology* 22, 108.

Papantonis, A., and Cook, P.R. (2013). Transcription Factories: Genome Organization and Gene Regulation. *Chem. Rev.* 113, 8683–8705.

Phillips, J.E., and Corces, V.G. (2009). CTCF: Master Weaver of the Genome. *Cell* 137, 1194–1211.

Pickersgill, H., Kalverda, B., de Wit, E., Talhout, W., Fornerod, M., and van Steensel, B. (2006). Characterization of the *Drosophila melanogaster* genome at the nuclear lamina. *Nat Genet* 38, 1005–1014.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–283.

Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K.C., Grüning, B.A., Villaveces, J., Habermann, B., Akhtar, A., and Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* 9, 189.

Rao, S.S.P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell* 171, 305-320.e24.

Redin, C., Brand, H., Collins, R.L., Kammin, T., Mitchell, E., Hodge, J.C., Hanscom, C., Pillalamarri, V., Seabra, C.M., Abbott, M.-A., et al. (2017). The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nature Genetics* 49, 36–45.

Redolfi, J., Zhan, Y., Valdes-Quezada, C., Kryzhanovska, M., Guerreiro, I., Iesmantavicius, V., Pollex, T., Grand, R.S., Mulugeta, E., Kind, J., et al. (2019). DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. *Nat Struct Mol Biol* 26, 471–480.

Rivera-Pomar, R., and Jäckle, H. (1996). From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. *Trends in Genetics* 12, 478–483.

Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. *Nat Rev Genet* 19, 789–800.

- Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V., and Furlong, E.E.M. (2007). A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev* *21*, 436–449.
- Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* *489*, 109–113.
- Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* *20*, 437–455.
- Schor, I.E., Bussotti, G., Maleš, M., Forneris, M., Viales, R.R., Enright, A.J., and Furlong, E.E.M. (2018). Non-coding RNA Expression, Function, and Variation during *Drosophila* Embryogenesis. *Curr Biol* *28*, 3547–3561.e9.
- Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H., Mirny, L., et al. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature* *551*, 51–56.
- Serfling, E., Jasin, M., and Schaffner, W. (1985). Enhancers and eukaryotic gene transcription. *Trends in Genetics* *1*, 224–230.
- Sexton, T., and Cavalli, G. (2015). The Role of Chromosome Domains in Shaping the Functional Genome. *Cell* *160*, 1049–1059.
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the *Drosophila* Genome. *Cell* *148*, 458–472.
- Sharifi-Zarchi, A., Gerovska, D., Adachi, K., Totonchi, M., Pezeshk, H., Taft, R.J., Schöler, H.R., Chitsaz, H., Sadeghi, M., Baharvand, H., et al. (2017). DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. *BMC Genomics* *18*, 964.
- Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* *488*, 116–120.
- Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* *15*, 272–286.
- Small, S., and Arnosti, D.N. (2020). Transcriptional Enhancers in *Drosophila*. *Genetics* *216*, 1–26.
- Spielmann, M., Brancati, F., Krawitz, P.M., Robinson, P.N., Ibrahim, D.M., Franke, M., Hecht, J., Lohan, S., Dathe, K., Nardone, A.M., et al. (2012). Homeotic Arm-to-Leg Transformation Associated with Genomic Rearrangements at the PITX1 Locus. *The American Journal of Human Genetics* *91*, 629–635.
- Spielmann, M., Lupiáñez, D.G., and Mundlos, S. (2018). Structural variation in the 3D genome. *Nat Rev Genet* *19*, 453–467.

- Stadhouders, R., van den Heuvel, A., Kolovos, P., Jorna, R., Leslie, K., Grosveld, F., and Soler, E. (2012). Transcription regulation by distal enhancers. *Transcription* 3, 181–186.
- Stadhouders, R., Fillion, G.J., and Graf, T. (2019). Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* 569, 345–354.
- Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* 24, 390–400.
- Symmons, O., Pan, L., Remeseiro, S., Aktas, T., Klein, F., Huber, W., and Spitz, F. (2016). The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. *Dev Cell* 39, 529–543.
- Szabo, Q., Bantignies, F., and Cavalli, G. (2019). Principles of genome folding into topologically associating domains. *Sci. Adv.* 5, eaaw1668.
- Szabo, Q., Cavalli, G., and Bantignies, F. (2021). Higher-Order Chromatin Organization Using 3D DNA Fluorescent In Situ Hybridization. In *Capturing Chromosome Conformation: Methods and Protocols*, B. Bodega, and C. Lanzuolo, eds. (New York, NY: Springer US), pp. 221–237.
- Thisse, C., Perrin-Schmitt, F., Stoetzel, C., and Thisse, B. (1991). Sequence-specific transactivation of the *Drosophila* twist gene by the dorsal gene product. *Cell* 65, 1191–1201.
- Thomas, S., Li, X.-Y., Sabo, P.J., Sandstrom, R., Thurman, R.E., Canfield, T.K., Giste, E., Fisher, W., Hammonds, A., Celniker, S.E., et al. (2011). Dynamic reprogramming of chromatin accessibility during *Drosophila* embryonic development. *Genome Biology* 12, R43.
- Tyndall, C., La Mantia, G., Thacker, C.M., Favaloro, J., and Kamen, R. (1981). A region of the polyoma virus genome between the replication origin and late protein coding sequences is required in cis for both early gene expression and viral DNA replication. *Nucleic Acids Res* 9, 6231–6250.
- Venken, K.J.T., Schulze, K.L., Haelterman, N.A., Pan, H., He, Y., Evans-Holm, M., Carlson, J.W., Levis, R.W., Spradling, A.C., Hoskins, R.A., et al. (2011). MiMIC: a highly versatile transposon insertion resource for engineering *Drosophila melanogaster* genes. *Nat Methods* 8, 737–743.
- Wit, E. de, and Laat, W. de (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* 26, 11–24.
- Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., and Romano, L.A. (2003). The Evolution of Transcriptional Regulation in Eukaryotes. *Molecular Biology and Evolution* 20, 1377–1419.
- Yin, Z., Xu, X.L., and Frasch, M. (1997). Regulation of the twist target gene tinman by modular cis-regulatory elements during early mesoderm development. *Development* 124, 4971–4982.
- Yokoshi, M., Segawa, K., and Fukaya, T. (2020). Visualizing the Role of Boundary Elements in Enhancer-Promoter Communication. *Molecular Cell* 78, 224–235.e5.



Zeitlinger, J. (2020). Seven myths of how transcription factors read the cis-regulatory code. *Current Opinion in Systems Biology* 23, 22–31.

Zhan, Y., Mariani, L., Barozzi, I., Schulz, E.G., Blüthgen, N., Stadler, M., Tiana, G., and Giorgetti, L. (2017). Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome Res* 27, 479–490.

Zhang, H., Emerson, D.J., Gilgenast, T.G., Titus, K.R., Lan, Y., Huang, P., Zhang, D., Wang, H., Keller, C.A., Giardine, B., et al. (2019). Chromatin structure dynamics during the mitosis-to-G1 phase transition. *Nature* 576, 158–162.

Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M., and Furlong, E.E.M. (2009). Combinatorial binding predicts spatio-temporal cis -regulatory activity. *Nature* 462, 65–70.

Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., et al. (2021). Nonlinear control of transcription through enhancer-promoter interactions. *BioRxiv* 2021.04.22.440891.